Marginalized Mixture Models for Zero-Inflated Longitudinal Count Data

PhD Thesis

Md. Ershadul Haque
Registration No.: 117/2018-2019
Department of Statistics
University of Dhaka, Bangladesh

June, 2023

Marginalized Mixture Models for Zero-Inflated Longitudinal Count Data

A Thesis submitted for the Degree of ***Doctor of Philosophy***

Submitted by

Md. Ershadul Haque
Registration No.: 117/2018-2019
Department of Statistics
University of Dhaka, Bangladesh

Supervised by

Wasimul Bari, PhD
Professor
Department of Statistics
University of Dhaka, Bangladesh

June, 2023

Acknowledgements

I would like to praise Allah, the Almighty and the Most Merciful for His blessing given to me to complete this thesis and for providing me with pathways whenever I faced any obstacles while conducting the study.

This is a great pleasure to express my deepest indebtedness and immense gratitude to my respected supervisor Prof. Dr. Wasimul Bari, Department of Statistics, University of Dhaka. His constant cooperation by offering me advice, encouragement, and guidance throughout the entire period of work ensured its successful completion.

It is very unfortunate and heartbreaking that I lost my co-supervisor Prof. Dr. Taslim Sazzad Mallick, Department of Statistics, University of Dhaka, during this work. He had guided me through many statistical and computing ideas at the initial stage of this thesis. I express my deepest gratitude and indebtedness to him and pray to the Almighty for the eternal peace for his departed soul.

Special thanks are to my respected teacher, Prof. Lutfor Rahaman, Former Chairman, Department of Statistics, University of Dhaka, for providing me his administrative support to enroll as a PhD student at the Department of Statistics, University of Dhaka. I am also thankful to Prof. Sayema Sharmin, Chairperson, Department of Statistics, University of Dhaka, for her encouragement and advice as well as administrative support throughout the entire period of work.

I offer special thanks to Prof. Dr. Jafar Ahmed Khan, Department of Statistics, University of Dhaka, for his continuous suggestions and encouragement for completing the work successfully. With a deep sense of gratitude, I also thank Dr. Zillur Rahman Shabuz, Associate Professor, Department of Statistics, University of Dhaka, for his continuous directions about LaTeX software for completing the dissertation. I would like to express my sincere gratitude to all the faculty members of the Department of Statistics, University of Dhaka, for their cooperation during the work.

I would like to express my indebtedness and heartfelt gratefulness to my wife Mrs. Roksana Afroz Munny, my two daughters Rimsha Eshal Aqsa and Tanisha Eshan Ayda, my beloved parents for their endless love, prayers and continuous support and encouragement.

I would also like to thank University Grants Commission (UGC), Bangladesh to offer me the UGC PhD fellowship to complete this research work.

Abstract

In practice, the count data may contain too many structures, which can cause the zero-augmentation issue. If such data are analyzed using standard count models, the results can be misleading. Traditionally, zero-inflated data are analyzed using a statistical model assuming that data arise from a standard count as well as a degenerated populations. Since zero-truncated count models provide similar results obtained from traditional zero-inflated count models, in this study, we have proposed a marginalized statistical model based on mixture of two-component Poisson distributions for analyzing zero-inflated longitudinal count data (clustered and repeated measures data) to draw inference regarding the effects of the covariates on marginal mean (marginalization over Poisson components) of the count response.

To analyze the zero-inflated clustered data, our proposed marginalized Poisson-Poisson (REMPois-Pois) mixture model takes into account the intra-cluster correlation by incorporating random effects into the models for marginal mean and component-1 mean in the existing marginalized Poisson-Poisson (MPois-Pois) mixture model suggested for cross-sectional setup. The parameters of the REMPois-Pois model were estimated using maximum likelihood (ML) technique. The Gauss–Hermite quadrature (GHQ) technique was employed to approximate the integrals appeared in the likelihood function. The performance of the proposed marginalized model were assessed through extensive simulation studies. It was observed that the proposed model performs well under different scenarios of simulation setups. Finally, the proposed REMPois-Pois model was illustrated by using a nationally represen-

tative data set on the number of antenatal care (ANC) visits extracted from Bangladesh Demographic and Health Survey (BDHS), 2014.

To analyze zero-inflated longitudinal repeated measures count data, a marginalized mixture of two-component longitudinal Poisson models (RMMPois-Pois model) have also been proposed in this study. Since observations obtained from the same subject are likely to be correlated in such instance, the regression parameters of the model were estimated by generalized quasilikelihood (GQL) approach taking true correlation into account. To examine the performance of the RMMPois-Pois model, we have conducted extensive simulation studies. The results of the simulation studies indicate that the performance of the proposed model is remarkable. To illustrate the RMMPois-Pois model, a real life repeated count data set on the number of episodes for certain side effect acquired from a pharmaceutical company was utilized.

Contents

Acknowledgements i

Abstract iii

1 Introduction 1

2 Marginalized Mixture Models: Cross-Sectional Setup 10

2.1 Marginalized Zero-Inflated Poisson Model 11

2.2 Poisson-Poisson Mixture Model 15

2.3 Marginalized Poisson-Poisson Mixture Model 18

2.3.1 Likelihood Function 19

2.3.2 Score Function 20

2.4 Simulation Study 22

2.5 Illustration 37

2.5.1 Data 37

2.5.2 Variables 37

2.5.3 Results 38

2.6 Conclusion 41

3 Marginalized Mixture Models: Zero-Inflated Clustered Count Data 43

3.1 Random Effects Marginalized Poisson-Poisson Mixture Model 45

3.1.1 Likelihood Function 47

3.1.2 Score Function 49

3.1.3 Intraclass Correlations 51

3.2 Simulation Study 53

3.2.1 The situation when $\rho^* = 0$ 54

3.2.2 The situation when $\rho^* > 0$ 70

3.3 Illustration 71

3.3.1 Results 71

3.4 Conclusion 75

4 Marginalized Mixture Models: Zero-Inflated Repeated Measures Count Data 77

4.1 MPois-Pois Model for Repeated Measures Count Data 81

4.1.1 Two-component Mixture of Longitudinal Poisson Models 81

4.1.2 Marginalized Mixture of Two-Component Longitudinal Poisson Models 83

4.2 Estimation: GLM in Longitudinal Setup 84

4.2.1 Generalized Estimating Equation Method 84

4.2.2 Generalized Quasi-Likelihood Method 86

4.3 GQL Estimation: RMMPois-Pois Model 87

4.3.1 Estimating Mixing Proportion 89

4.3.2 Estimation of Regression Parameters: Marginal Model 92

4.3.3 Estimation of Correlation Parameter: Marginal Model 93

4.3.4 Estimating Regression Parameters: Component-1 Model 94

4.3.5 Estimation of Correlation Parameter: Component-1 Model 95

4.4 Simulation Study 96

4.4.1 Marginalized Poisson AR(1)-Poisson AR(1) Mixture Probability Model 96

4.4.2 Marginalized Poisson MA(1)-Poisson MA(1) Mixture Probability Model 101

4.4.3 Marginalized Poisson EQCOR - Poisson EQCOR Mixture Probability Model 104

4.4.4 Performance of the Proposed RMMPois-Pois Model 106

4.5 Illustration 119

4.6 Conclusion 121

5 Conclusions **122**

Appendices **127**

A **128**

A.1 Hessian: MPois-Pois Model 128

A.2 The EM Algorithm: Poisson-Poisson Mixture Model 130

B **133**

B.1 Gauss-Hermite numerical quadrature for evaluating integrals 133

B.2 REMPois-Pois Model: Score function 135

B.3 The EM Algorithm: Random Effects Poisson-Poisson Mixture Model 137

B.4 Simulation Results: Intra-cluster Correlation Coefficient from the REMPois-Pois Model 143

C **145**

C.1 Simulation Results: Marginalized Mixture of Poisson AR(1) Model 145

C.2 Simulation Results: Marginalized Mixture of Poisson MA(1) Model 149

C.3 Simulation Results: Marginalized Mixture of Poisson EQCOR Model 153

C.4 Simulation Results: Marginalized Mixture of Longitudinal Poisson Model . . 157

List of Figures

2.1 A comparison of MSE estimated from MPois-Pois, MZIP, negative binomial (NB) and Poisson (Pois) models with varying sample sizes for regression parameter β_0 34

2.2 A comparison of MSE estimated from MPois-Pois, MZIP, negative binomial (NB) and Poisson (Pois) models with varying sample sizes for regression parameter β_1 35

2.3 A comparison of MSE estimated from MPois-Pois, MZIP, negative binomial (NB) and Poisson (Pois) models with varying sample sizes for regression parameter β_2 36

2.4 Distribution of number of ANC visits during pregnancy, BDHS 2014 39

3.1 A comparison of MSE from REMPois-Pois models with different mixing probability and varying cluster sizes for the regression parameter $\beta_0 = 0.20$ 66

3.2 A comparison of MSE from REMPois-Pois models with different mixing probability and varying cluster sizes for the regression parameter $\beta_1 = 0.40$ 67

3.3 A comparison of MSE from REMPois-Pois models with different mixing probability and varying cluster sizes for the regression parameter $\beta_2 = 0.30$ 68

3.4 A comparison of MSE from REMPois-Pois models with different mixing probability and varying cluster sizes for the regression parameter $\sigma_u = 0.40$ . . . 69

4.1 A comparison of MSE from RMMPois-Pois models with different mixing probability and varying number of individuals for the regression parameter $\beta_0 = 0.20$ 115

4.2 A comparison of MSE from RMMPois-Pois models with different mixing probability and varying number of individuals for the regression parameter $\beta_1 = 0.60$ 116

4.3 A comparison of MSE from RMMPois-Pois models with different mixing probability and varying number of individuals for the regression parameter $\beta_2 = 0.50$ 117

4.4 A comparison of Bias from RMMPois-Pois models with different mixing probability and varying number of individuals for the lag-1 correlation ρ_1 118

List of Tables

2.1 Simulated mean (SM), simulated mean of mixing proportion (SMP), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr.) for estimating Marginal Parameters ($\boldsymbol{\beta}$); Component-1 Parameters ($\boldsymbol{\alpha}$) and τ under MPois-Pois model for MPois-Pois dgp 26

2.2 Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr.) for estimating Marginal Parameters: ($\boldsymbol{\beta}$) under MZIP, NB, Pois, ZIP and PH models along with Component-1 Parameters ($\boldsymbol{\alpha}$) for MZIP model for MPois-Pois dgp . . . 28

2.3 The estimated marginal parameters ($\hat{\beta}$), standard errors (SE), and p-values under MPois-Pois mixture model on number of ANC visits, BDHS 2014 . . . 40

3.1 Simulated mean (SM), simulated mean of mixing proportion (SMP), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) for estimating Marginal Parameters: ($\boldsymbol{\beta}, \sigma_u$); Component-1 Parameters ($\boldsymbol{\alpha}, \sigma_c$) and τ for mixing proportion $\pi^* = 0.50$ from REMPois-Pois model for different number of clusters and various cluster sizes 58

3.2 Simulated mean (SM), simulated mean of mixing proportion (SMP), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) for estimating Marginal Parameters: $(\boldsymbol{\beta}, \sigma_u)$; Component-1 Parameters $(\boldsymbol{\alpha}, \sigma_c)$ and τ for mixing proportion $\pi^* = 0.70$ from REMPois-Pois model for different number of clusters and various cluster sizes 60

3.3 Simulated mean (SM), simulated mean of mixing proportion (SMP), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) for estimating Marginal Parameters: $(\boldsymbol{\beta}, \sigma_u)$; Component-1 Parameters $(\boldsymbol{\alpha}, \sigma_c)$ and τ for mixing proportion $\pi^* = 0.90$ from REMPois-Pois model for different number of clusters and various cluster sizes 63

3.4 Results from simulation study for estimating marginal parameters: $(\boldsymbol{\beta}, \sigma_u)$; component-1 parameters $(\boldsymbol{\alpha}, \sigma_c)$ from REMPois-Pois model by considering correlated $\boldsymbol{u}$ and $\boldsymbol{c}$ (random effects for marginal and subpopulation) with mixing proportion $\pi^* = 0.70$ and cluster size 10 for 200 clusters 71

3.5 The estimated parameters and other statistics for REMPois-Pois models for analyzing the number of ANC visits during a pregnancy period, BDHS 2014 72

4.1 Simulated mean (SM), Bias, estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters $(\boldsymbol{\beta})$; and component-1 parameters $(\boldsymbol{\alpha})$ with mixing proportion $\pi^* = 0.70$ from marginalized mixture of Poisson AR(1) model for different values of K and T 107

4.2 Simulated mean (SM), Bias, estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.70$ from marginalized mixture of Poisson MA(1) model for different values of K and T 109

4.3 Simulated mean (SM), Bias, estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.70$ from marginalized mixture of Poisson EQCOR model for different values of K and T 111

4.4 RMMPois-Pois Model estimates of parameters, standard errors (SE), and the IRR for analyzing repeated number of episodes of a certain side effect 120

4.5 Lag-l correlation estimated from the proposed RMMPois-Pois model 121

B.1 The ICC and the estimated ICC from REMPois-Pois model for different number of clusters and various cluster sizes 143

C.1 Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.50$ from marginalized mixture of Poisson AR(1) model for $\rho^* = 0.40$ and for different values of K and T 145

C.2 Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.90$ from marginalized mixture of Poisson AR(1) model for $\rho^* = 0.40$ and for different values of K and T 147

C.3 Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.50$ from marginalized mixture of Poisson MA(1) model for $\rho^* = 0.40$ and for different values of K and T 149

C.4 Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.90$ from marginalized mixture of Poisson MA(1) model for $\rho^* = 0.40$ and for different values of K and T 151

C.5 Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.50$ from marginalized mixture of Poisson EQCOR model for $\rho^* = 0.40$ and for different values of K and T 153

C.6 Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.90$ from marginalized mixture of Poisson EQCOR model for $\rho^* = 0.40$ and for different values of K and T 155

C.7 Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.50$ from marginalized mixture of Poisson AR(1), Poisson MA(1), and Poisson EQCOR model for $\rho^* = 0.70$ and $T = 4$, and for different values of K 157

C.8 Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.70$ from marginalized mixture of Poisson AR(1), Poisson MA(1), and Poisson EQCOR model for $\rho^* = 0.70$ and $T = 4$, and for different values of K 159

C.9 Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.90$ from marginalized mixture of Poisson AR(1), Poisson MA(1), and Poisson EQCOR model for $\rho^* = 0.70$ and $T = 4$, and for different values of K 161

Chapter 1

Introduction

Poisson regression is considered as a foundation for modeling different types of count data appeared in many field of applied researches. However, application of standard Poisson model is based on an assumption of equality in the mean and variance of the values of outcome variable.

In some situations, the observed count data consist of too many structures that may lead to the problem of zero-augmentation as well as overdispersion (Benecha et al., 2017). Analysis of such data using standard count models may provide misleading conclusions (Frühwirth-Schnatter, 2006; Wedel and DeSarbo, 1995; Wang, 1994; Wang et al., 1996). To overcome this problem, the mixture of Poisson distributions has been suggested instead of one-component Poisson distribution (Wang et al., 1996) as the mixture model framework provides a model for coping with local variation in the data (McLachlan and Peel, 2000). To analyze heterogeneous data, Dempster et al. (1977) first proposed finite mixture distributions. The term 'unobserved heterogeneity' will be used in this dissertation to express the type of heterogeneity in the data that cannot be controlled by imposing all the known covaritaes into the model. To address unobserved heterogeneity, Wang et al. (1996) had fitted a mixture of Poisson regression models for analyzing count data with many structures.

Count data with excess zero arises in many areas such as engineering, biomedical, public health, demography, economics, and social science (Mullahy, 1986; Lambert, 1992; Böhning et al., 1999; Shankar et al., 1997; Porter et al., 2012; Hu et al., 2011). Since the standard Poisson regression model-based analysis of such count data yields misleading inference about the parameters of interest due to the fact that the observed proportion of zero counts is much higher than expected under the fitted Poisson model, Mullahy (1986) and Lambert (1992) recommended using a different distribution for the zeros in a mixture model framework in order to make valid inferences in modeling the zero-inflated count data.

The general framework of existing count models with excess zeros are based on mixture of two subpopulations classified as 'at-risk' (or susceptible) and 'not-at-risk' groups. Using a degenerate distribution with mass 1 for the zero outcome, the subpopulation designated as 'not-at-risk' generates structural zeros. The subpopulation defined by 'at-risk' group may provide non-negative counts following a count distribution or positive counts following a zero truncated count distribution. Depending on the data generating process (dgp) for counts in 'at-risk' group, hurdle or zero inflated count model is used in the literature where zero heavy Poisson count data are assumed to be generated either following a Zero-Inflated Poisson (ZIP) model (Lambert, 1992; Hu et al., 2011; Minami et al., 2007; Yip and Yau, 2005; Rose et al., 2006; Gurmu and Trivedi, 1996; Shonkwiler and Shaw, 1996) or a Poisson Hurdle (PH) model (Mullahy, 1986; Porter et al., 2012; Hu et al., 2011; Bilgic and Florkowski, 2007; Rose et al., 2006; Pohlmeier, 1996; Welsh et al., 1996; Gurmu, 1997; Gurmu and Trivedi, 1996). Note that for ZIP model it is possible for susceptible classes to provide zero counts as well, these are known as sampling zeros. An overview of zero-inflated count data models including types of zeros ('sampling zeros' and 'structural zeros') is available in literatures (Cameron and Trivedi, 2013; Hilbe, 2014) and the references therein.

Traditionally, the inference procedure for the zero-inflated model is based on fitting mean parameters of the Poisson distribution adjusting the excess zeros arising from the binary dgp. The adjustment of excess zeros requires separate independent binary response modelling approach (Lambert, 1992). Haque et al. (2022b) argued that since the mean of the parent Poisson distribution is the parameter of interest in both the Poisson and Zero-Truncated Poisson (ZTP) models, the ZTP model can be used as an alternative approach for analyzing zero-inflated count data under the framework of the ZIP model.

The existing mechanism for generating zero-augmented counts is a two-component mixture where one of the components forms a degenerate distribution at zero. A finite mixture model (FMM) of this type with a degenerate component is sometimes referred to as a non-standard mixture model (McLachlan and Peel, 2000). However, it may happen in practice that count data observations in a sample arise from two or more populations, and thus the resultant data for the whole population may contain zero-inflation and/or overdispersion (Wang et al., 1996; Benecha et al., 2017). Although the problem of overdispersion can be addressed by utilizing a 'negative binomial' model or zero-inflated negative binomial (ZINB) model (Greene, 1994), it is difficult to handle overdispersion in a situation when it arises due to a mixture of standard count distributions (Benecha et al., 2017). Moreover, a *non-standard mixture model* setup based on a degenerate latent class may not be appropriate or may provide a misleading conclusion in a setup when zero-augmentation arise from mixture of latent standard count distributions (Benecha et al., 2017). To overcome this difficulty, a *mixture of standard count models* can be applied to address different unobserved structures in the populations (Wang et al., 1996; Benecha et al., 2017).

This study aimed to analyze zero-inflated data arising from mixture of two 'at-risk' populations by incorporating a mixture of two standard count models. To develop the model, we assume that some counts originate from Poisson distributions with very low means and some from another Poisson with larger means. Therefore, we limit the dgp of zero heavy counts to a two-component Poisson mixture of a heterogeneous population known as *Poisson-Poisson mixture* distribution.

For instance, consider the number of ANC visits as an outcome variable of interest. The study population for analyzing the data regarding such an outcome variable consists of all women in the reproductive age group who have completed at least one pregnancy in their lifetime. In this situation, all women have a positive probability of providing a non-negative count and hence there is no counts expected from a degenerate zero population in the data. Data collected from this population may result in zero-inflated data. This may be due to the study population consists of two susceptible classes one with very low means and another with larger means. From the available literature (Haque et al., 2022b; Bekalo and Kebede, 2021; Bhowmik et al., 2020; Afolabi and Agbaje, 2018), it can be found that the researchers frequently use existing zero-augmented count models (ZIP, PH, ZTP, and ZINB) to analyze such data. In such situations, they assumed some zeros were from the degenerate population. In this instance, they assume some zeros were obtained from some women other than those in the reproductive age group. This assumption is not appropriate in general and resulted in a flaw in the study population. To overcome such flaw in the study population, following Wang et al. (1996), the current study suggest to analyze such data buy using a two-component mixture of Poisson distribution.

Because of latent class formulation, the interpretations of parameters in terms of incidence rate ratio (IRR) under mixture model setup are often imprecise or misleading. To solve the difficulty, the significance of *marginal inference (marginalization over the subpopulations)* under the framework of mixture models is well documented in many studies (Albert et al., 2014; Böhning et al., 1999; Preisser et al., 2012; Long et al., 2014; Benecha et al., 2017). The objective in such instances was to estimate the exposure effects on the entire population. In mixture model setup, a marginal mean is defined by the overall mean obtained by averaging the latent mean response across the distribution of the susceptibility status, regardless of covariates. However, under the mixture model setup, it is not possible to directly infer the marginal mean from the latent class inference of parameters (Albert et al., 2014; Preisser et al., 2012; Benecha et al., 2017). Long et al. (2014), Albert et al. (2014) and Preisser et al. (2012) made some efforts to overcome the difficulties by considering marginalized mixture modeling approach under existing non-standard mixture model for zero-inflated count data in cross-sectional settings. Furthermore, Benecha et al. (2017) proposed a marginalized mixture model for analyzing zero-inflated count data originated from two 'at-risk' populations in cross-sectional settings in which the marginal parameters and the nuisance parameters can be estimated by using the maximum likelihood (ML) method of estimation. Since the aim of this study is to facilitate the interpretation of regression parameters in terms of IRR, this study focuses inference procedure based on the marginal mean of two-component Poisson mixture distribution.

The analysis of *zero-inflated longitudinal count data* using existing zero-inflated models has recently drawn a lot of attention from researchers. In longitudinal data, observations obtained from the same observation unit are likely to be correlated. One should take this correlation into account in developing methods for such data. One can view the longitudinal data as clustered data or repeated measures data. In clustered data, the sampling unit is a group of observation units, whereas in repeated measurements, the sampling unit is an observation unit where more than one observations are taken from the same observation units at different occasions. Moreover, many studies in applied statistics utilize data that are not perfectly longitudinal with reference to temporal order but can be treated as clustered. For example, when intact groups are randomized to health interventions or naturally occurring groups in the population are randomly sampled, this can be considered clustered data. It is reasonable to expect that measurements on units within a cluster are more similar than measurements on units in different clusters. The degree of clustering can be expressed in terms of correlation among the measurement units within the same cluster (Fitzmaurice et al., 2012). The distinctive feature of the observations within a cluster is that they exhibit positive correlation.

Hall (2000) suggested a ZIP model modification for analyzing zero-augmented clustered count data that includes random effects in the Poisson process to take intraclass correlation into consideration. In order to account for excess zeros as well as over-dispersion in correlated data, Yau et al. (2003) proposed a zero-inflated negative binomial (ZINB) regression model with independent random effects in each process. As a two-part model, hurdle have also been applied for clustered count data (Min and Agresti, 2005).

In some clinical contexts, as was previously noted, it is preferable to draw conclusions from the marginal mean rather than the means of the two latent classes. Characterizing marginal means and associated marginal effects of covariates is usually difficult when model parameters are estimated under existing non-standard mixture modeling approach of *zero-inflation for clustered count data*, especially when both portions of the model have random effects (Su et al., 2015; Tom et al., 2016; Long et al., 2015). For taking care of this difficulty, the researchers have developed marginalized models in case of existing non-standard mixture modeling approach of zero-inflation for clustered count data (Tabb et al., 2016; Long et al., 2015; Kassahun et al., 2014; Lee et al., 2011). In the context of longitudinal semi-continuous data, Farewell et al. (2017) have highlighted some of the difficulties associated with drawing marginal inferences from two-part models. An excellent discussion of various zero-inflated count models for longitudinal data and inference regarding their marginal means is available in Farewell et al. (2017) and the references therein.

To analyze *repeated measures zero-inflated count data*, Hasan and Sneddon (2009) proposed an observation driven ZIP models as an extension of the cross-sectional ZIP model and Hall and Zhang (2004) proposed GEE based marginal (marginalization over the observation units) ZIP model. Also a comparative study for fitting observation driven ZIP models and parameter driven ZIP models have been conducted by Hasan et al. (2016) if the count data were collected repeatedly over time. The marginal inference (marginalization over the subpopulations) from the non-standard mixture (ZIP) model had not studied yet to analyze repeated measures zero-inflated count data. But we are restrained from developing marginal models for such a mechanism of zero-inflation because we have already mentioned that this mechanism is inappropriate when data are arising from 'at-risk' populations.

Although Benecha et al. (2017) proposed marginally-specified mean models for mixtures of two count distributions under a cross-sectional setup, the model still requires some challenges for further extension in the analysis of zero-inflated longitudinal (clustered and repeated measures) count data. Considering these challenges, an extension for these models using two-component Poisson mixture is proposed in this study along with its inference procedure.

Following Benecha et al. (2017), we propose a marginalized Poisson-Poisson mixture model in the current study to analyzing zero-inflated clustered count data and zero-inflated repeated measures count data. Following McKenzie (1988), Sutradhar (2003) has generated repeated measures Poisson count data using observation-driven models by employing some appropriate stationary AR(1), MA(1) and exchangeable autocorrelation structures. To conduct extensive simulation studies for the dgp under the proposed zero-inflated repeated measures count model, two-component mixture of observation-driven Poisson model (Sutradhar, 2003) has been used.

In order to propose marginalized models from mixture of 'at-risk' populations for zero-inflated longitudinal count data, the specific objectives of this dissertation are

- to propose a marginalized Poisson-Poisson mixture model for analyzing
  - i) zero-inflated clustered count data
  - ii) zero-inflated repeated measures count data;
- to develop the inference process for the proposed models for clustered data and repeated measures data;
- to examine the performance of the proposed models by carrying out extensive simulation studies.
- to analyze real data sets using the proposed models for clustered data and repeated measures data.

A comprehensive study of the cross-sectional marginalized Poisson-Poisson (MPois-Pois) mixture model (Benecha et al., 2017) has been conducted in Chapter 2 with a view to extending it to the longitudinal setup. This chapter describes some fundamental mixture modeling techniques for analyzing zero-inflated data, including reviews of the MZIP and MPois-Pois models. Comparison of these model along with other zero-inflated count models has been conducted through extensive simulation studies. The MPois-Pois model has been illustrated by utilizing a real data set in cross-sectional context. In Chapter 3, we have developed the random effects marginalized Poisson-Poisson (REMPois-Pois) mixture model for analyzing zero-inflated clustered count data. The method of handling the integration with respect to the random effects while maximizing the likelihood function is described and the performance of the proposed model are examined through extensive simulation studies. At the end of Chapter 3, a nationally representative clustered data set extracted from Bangladesh Demographic and Health Survey (BDHS), 2014 is used to illustrate the proposed model. In Chapter 4, we have developed the repeated measures marginalized Poisson-Poisson (RMMPois-Pois) mixture model (marginalization is considered over the subpopulations) for analyzing zero-inflated repeated measures count data. The GQL estimation method has been utilized for estimating the regression parameters of the proposed model and correlation parameters have been estimated by method of moments. The performance of the proposed model is examined through extensive simulation studies. A real data set is used to illustrate the proposed model for zero-inflated repeated measures count data. Chapter 5 concludes this study with overall findings and recommendations for future research.

Chapter 2

Marginalized Mixture Models: Cross-Sectional Setup

The mixture modeling approach are frequently used in modeling zero-augmented count data in the literature. The significance of inferences regarding the marginal mean under the framework of FMM has been well demonstrated in several research (Albert et al., 2014; Böhning et al., 1999; Preisser et al., 2012; Long et al., 2014). The objective of the researchers in such instances was to estimate the exposure effects on the entire population mean, i.e., the marginalization of means over the subpopulations.

Zero-inflated (ZI) models have become popular in analysing data arising from many areas of research such as biomedical, public health, engineering, ecology, demography, economics, and social science over the past two decades. The existing ZI models such as ZIP and PH are formulated using mixture model mechanism by considering an extra mass at the point zero. Despite their increasing popularity, some researchers has been pointed out the shortcoming of these models because of their latent class formulation where the mean response of the so-called 'at-risk' or susceptible population and the susceptibility probability are both related to covariates. The interpretations of parameters in terms of IRR are often imprecise or misleading. Particularly, it fails to discriminate between inference for the class of 'at-risk' population and inference for the overall exposure effects. To overcome such difficulty, Long et al. (2014) proposed a marginally specified mean model for analysing zero inflated count data in the framework of ZIP model. Following Long et al. (2014), marginalized mixture models of two 'at-risk' classes have been proposed by Benecha et al. (2017).

In this chapter, we have considered a marginalized Poisson-Poisson mixture (MPois-Pois) model for analyzing zero-augmented count data in cross-sectional settings. It is considered as a two-component Poisson mixture model (standard mixture) developed for obtaining marginal inference of the parameters. A comparison of MPois-Pois model has been made with the existing count models by conducting extensive simulation studies. The existing count models include Poisson model as a standard count model, negative binomial model in the presence of over-dispersion, and the two components non-standard mixture using a Poisson model along with its marginalized model in case of zero-augmentation. In this instance, we will limit our comparison to the PH model in the hurdle model specification, ZIP model in the ZI model specification.

We will explore the marginalized ZIP (MZIP) model and MPois-Pois model for analyzing zero-inflated and/or over-dispersed count data in a cross-sectional setup with inferential procedures. Finally, an attempt has been made to find out the potential determinants of the number of ANC visits taken by women during pregnancy period by using a nationally representative data extracted from Bangladesh Demographic and Health Survey (BDHS).

2.1 Marginalized Zero-Inflated Poisson Model

The ZI count model is considered as a non-standard two components mixture model. In ZI model, the first component models the probability that zero count arises from 'not-at-risk' subpopulation i.e., zero follows a degenerated distribution. If the observation does not follow a degenerated distribution at zero, the second component models the counts, including the 'at-risk' zeros, arise from a standard count distribution. The pmf of ZI model is specified as

$$\begin{aligned} Pr[Y_i = 0] &= f_1(0) + \big[1 - f_1(0)\big] f_2(0) \\ Pr[Y_i = s] &= \big[1 - f_1(0)\big] f_2(y_i), \qquad s = 1, 2, \ldots. \end{aligned} \tag{2.1}$$

In equation Eq.(2.1), the mixing probability $f_1(0)$ is associated with the degenerated distribution and $f_2(\cdot)$ is the probability mass function (pmf) of the standard count distribution. If counts for susceptible class arise following a Poisson distribution, the resulting model is called Zero-Inflated Poisson (ZIP). In ZIP model, the response variables $Y_i, i = 1, ..., n$ are independent; $Y_i = 0$ from degenerated population with probability ψ_i and $Y_i \sim \text{Poisson}(\lambda_i)$ with probability $(1 - \psi_i)$. The ZIP model is then obtained from Eq.(2.1) as

$$\begin{aligned} Pr[Y_i = 0] &= \psi_i + (1 - \psi_i) e^{-\lambda_i} \\ Pr[Y_i = s] &= \frac{(1 - \psi_i) e^{-\lambda_i} \lambda_i^s}{s!}, \qquad s = 1, 2, \ldots. \end{aligned} \tag{2.2}$$

The mean and standard deviation of ZIP random variable are obtained from Eq.(2.2) as

$$\mu_i^{\text{ZIP}} = (1 - \psi_i)\lambda_i$$

$$\sigma_i^{\text{ZIP}} = \sqrt{\mu_i^{\text{ZIP}} + \Big(\frac{\psi_i}{1 - \psi_i}\Big) \mu_i^{\text{ZIP}\,2}}.$$

Following Lambert (1992), covariates can be introduced in the ZIP model with the logit and the log-link function as

$$\begin{aligned} \text{logit}\big(\psi_i\big) &= \log\Big(\frac{\psi_i}{1 - \psi_i}\Big) = \boldsymbol{z_i'\alpha}, \\ \log\big(\lambda_i\big) &= \boldsymbol{x_i'\gamma}. \end{aligned} \tag{2.3}$$

Note that if the ith subject is from 'not-at-risk' group, $\boldsymbol{z_i}$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_{p_1})'$ are $p_1 \times 1$ vector of covariates and parameters, respectively and if the ith subject is from susceptible class, $\boldsymbol{x_i}$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_{p_2})'$ are $p_2 \times 1$ vector of covariates and parameters, respectively. It can be permissible to use same covariates for both classes (i.e., $\boldsymbol{x_i} = \boldsymbol{z_i}$) with $p_1 = p_2 = p$.

Thus, the log-likelihood function of ZIP model may be of the form

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}) = \sum_{y_i=0} \left(\log \left(e^{\boldsymbol{x}_i' \boldsymbol{\alpha}} + e^{-\exp(\boldsymbol{x}_i' \boldsymbol{\gamma})} \right) \right) + \sum_{y_i>0} \left(y_i \boldsymbol{x}_i' \boldsymbol{\gamma} - e^{\boldsymbol{x}_i' \boldsymbol{\gamma}} \right) - \sum_{i=1}^{n} \left(\log \left(1 + e^{\boldsymbol{x}_i' \boldsymbol{\alpha}} \right) \right) - \sum_{y_i>0} \left(\log(y_i!) \right). \tag{2.4}$$

The parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ of ZIP model in Eq.(2.3) have latent class interpretations; that is, for one unit change in the kth element of $\boldsymbol{x}_i$, the odds of being in the degenerate population can be expressed in terms of α_k and the mean change in the population modeled by Poisson distribution can be expressed in terms of γ_k.

To obtain the exposure effect on the overall population mean directly from the ZIP model, let us consider $\mu_i \equiv E(Y_i)$ is the marginal mean. The relationship between μ_i and the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ from the ZIP model is

$$\mu_i = (1 - \psi_i)\lambda_i = \frac{e^{\boldsymbol{x}_i' \boldsymbol{\gamma}}}{1 + e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}}. \tag{2.5}$$

In Eq.(2.5), the overall population mean is a function of all covariates and parameters from both model parts. Thus, for the kth covariate, the ratio of means for a one-unit increase in x_{ik} is

$$\frac{\mu_i(x_{ik} = a + 1, \tilde{x}_i = \tilde{x}_i)}{\mu_i(x_{ik} = a, \tilde{x}_i = \tilde{x}_i)} = \exp(\gamma_k) \frac{1 + \exp\left(a\alpha_k + \tilde{x}_i' \tilde{\boldsymbol{\alpha}} \right)}{1 + \exp\left((a+1)\alpha_k + \tilde{x}_i' \tilde{\boldsymbol{\alpha}} \right)}, \tag{2.6}$$

where $\tilde{x}_i$ indicates all covariates except x_{ik} and $\tilde{\boldsymbol{\alpha}}$ is created by removing α_k from $\boldsymbol{\alpha}$. Thus, unless $\tilde{\boldsymbol{\alpha}} = \mathbf{0}$, the IRR is not constant across various levels of the extraneous covariates included in the ZIP model. Additionally, in order to make statements regarding the variability of any IRR estimates at fixed levels of the non-exposure covariates, formal statistical techniques, such as the delta method or bootstrap resampling methods are required. However, sophisticated computational methods are required to use these techniques in many applied analytics.

To overcome the difficulties in interpreting the IRR given in Eq.(2.6), Long et al. (2014) have suggested to model directly the overall mean μ_i to interpret the parameter in terms of overall exposure effects in the ZIP model. The suggested model is known as marginalized zero-inflated Poisson (MZIP) model. To model the marginal mean, MZIP model can be specified as

$$\begin{aligned} \operatorname{logit}\big(\psi_i\big) &= \boldsymbol{x}_i'\boldsymbol{\alpha}, \\ \log(\mu_i) &= \boldsymbol{x}_i'\boldsymbol{\beta}. \end{aligned} \tag{2.7}$$

Then, $\mu_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$ and thus, the effect of covariates can be explained in terms of IRR. Similar with ZIP model, it can be permissible to use same covariates for both the models for simplicity as shown in Eq.(2.7). Under MZIP, the log-likelihood function can be obtained by replacing λ_i in Eq.(2.4) in terms of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ as

$$\lambda_i = \frac{\mu_i}{1-\psi_i} = \exp(\delta_i). \tag{2.8}$$

where $\delta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \log\Big[1 + \exp\big(\boldsymbol{x}_i'\boldsymbol{\alpha}\big)\Big]$. Then the log-likelihood function of MZIP model can be expressed as

$$\begin{aligned} \ell(\boldsymbol{\alpha},\boldsymbol{\beta}) = &-\sum_{i=1}^{n}\left(\log\left(1+e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}\right)\right) + \sum_{y_i=0}\left(\log\left(e^{\boldsymbol{x}_i'\boldsymbol{\alpha}} + e^{-\exp(\boldsymbol{x}_i'\boldsymbol{\beta})(1+e^{\boldsymbol{x}_i'\boldsymbol{\alpha}})}\right)\right) \\ &+ \sum_{y_i>0}\left(-\left(1+e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}\right)e^{\boldsymbol{x}_i'\boldsymbol{\beta}} + y_i\log\left((1+e^{\boldsymbol{x}_i'\boldsymbol{\alpha}})\right) + y_i\boldsymbol{x}_i'\boldsymbol{\beta} - \log(y_i!)\right). \end{aligned} \tag{2.9}$$

The quasi-Newton optimization method can be implemented to obtain the estimates of the parameters of Eq.(2.9) (Long et al., 2014).

2.2 Poisson-Poisson Mixture Model

The Poisson mixture model is a popular standard FMM in analysing heterogeneous and/or over-dispersed count data, where the gth component of the FMM has a Poisson mass as

$$f_g(y_i; \mu_{g,i}) = \frac{e^{-\mu_{g,i}} \mu_{g,i}^{y_i}}{y_i!}, \tag{2.10}$$

where $\mu_{g,i}$ for $g = 1, ..., k$ is the mean of the ith response conditional on its membership on the gth component of the mixture. The component means are then modeled as a function of covariates through the link function as

$$\log(\mu_{g,i}) = \boldsymbol{x}'_{g,i} \boldsymbol{\beta}_g, \tag{2.11}$$

where $\boldsymbol{x}_{g,i}$ and $\boldsymbol{\beta}_g$ are the $p_g \times 1$ vector of covariates and parameters, respectively for the ith subject. When a mixture distribution is physically identifiable and the mixing proportions are known, subgroup analysis can be applied to the data. However, there are also many situations where the components cannot be identified with externally existing groups since the groups are treated as latent and it is apparently impossible to model the situation by a standard probability model.

In order to model the counts from heterogeneous populations where the components cannot be identified with externally existing groups, it is reasonable to assume that the counts are generated from a mixture of finite number of latent components (Wang et al., 1996; Benecha et al., 2017). Therefore in mixture model setup, linear function of separate latent class regression parameters are specified for the mean of each component of the mixture.

Suppose that in a mixture distribution there are two components (subpopulations). The count response, Y_i for the set of covariates of ith subject may come from one of the two components following a standard count distribution, i.e., $Y_i|G = g \sim f_g(y_i; \mu_{g,i}), g = 1, 2$ for $i = 1, ..., n$ with $Pr(G = 1) = \pi^*$ and $Pr(G = 2) = 1 - \pi^*$. Then, the marginal pmf of two components mixture distribution (Wang et al., 1996) can be derived as

$$\begin{aligned} Pr(Y_i = y_i) &= Pr(Y_i = y_i|G = 1) \times Pr(G = 1) + Pr(Y_i = y_i|G = 2) \times Pr(G = 2) \\ &= \pi^* f_1(y_i; \mu_{1,i}) + (1 - \pi^*) f_2(y_i; \mu_{2,i}). \end{aligned} \tag{2.12}$$

The pmf of the Poisson-Poisson mixture distribution is then obtained from Eq.(2.10) and Eq.(2.12) as follows

$$f(Y_i = y_i; \pi^*, \mu_{1,i}, \mu_{2,i}) = \pi^* \frac{e^{-\mu_{1,i}} \mu_{1,i}^{y_i}}{y_i!} + (1 - \pi^*) \frac{e^{-\mu_{2,i}} \mu_{2,i}^{y_i}}{y_i!}. \tag{2.13}$$

To study the properties of the mixture distribution it is convenient to express Eq.(2.13) in a hierarchy as follows

$$Y_i|d_i \sim \text{Pois}\big(d_i \mu_{1,i} + [1 - d_i] \mu_{2,i}\big), d_i = 0, 1, \tag{2.14}$$

where d_i is the realization of Bernoulli random variable D_i with $P[D_i = 1] = \pi^*$. Therefore the conditional mean and variance are as follows

$$E\Big[Y_i|d_i\Big] = d_i \mu_{1,i} + (1 - d_i) \mu_{2,i} = \text{Var}\Big[Y_i|d_i\Big].$$

Then the marginal mean and variance of the Poisson-Poisson mixture distribution are respectively computed as

$$E(Y_i) = E\big[E[Y_i|D_i]\big] = E\big[D_i\mu_{1,i} + (1-D_i)\mu_{2,i}\big] = \pi^*\mu_{1,i} + (1-\pi^*)\mu_{2,i} = \mu_i, \tag{2.15}$$

and

$$\begin{aligned}
\text{Var}(Y_i) &= E\big[\text{Var}(Y_i|D_i)\big] + \text{Var}\big[E[Y_i|D_i]\big] \\
&= E\big[D_i\mu_{1,i} + (1-D_i)\mu_{2,i}\big] + \text{Var}\big[D_i\mu_{1,i} + (1-D_i)\mu_{2,i}\big] \\
&= \big[\pi^*\mu_{1,i} + (1-\pi^*)\mu_{2,i}\big] + \big[\mu_{1,i}^2\text{Var}(D_i) + \mu_{2,i}^2\text{Var}(1-D_i) + 2\mu_{1,i}\mu_{2,i}\text{Cov}(D_i, 1-D_i)\big] \\
&= \mu_i + \Big[\mu_{1,i}^2\pi^*(1-\pi^*) + \mu_{2,i}^2(1-\pi^*)\pi^* + 2\mu_{1,i}\mu_{2,i}\big(-\pi^*(1-\pi^*)\big)\Big] \\
&= \mu_i + \pi^*(1-\pi^*)(\mu_{2,i} - \mu_{1,i})^2.
\end{aligned} \tag{2.16}$$

Thus for a Poisson-Poisson mixture distribution, $\text{Var}(Y_i) > E(Y_i)$ unless $\mu_{1,i} = \mu_{2,i}$. It implies that the Poisson-Poisson mixture distribution can be used for modeling overdispersed count data. In Poisson-Poisson mixture regression model, the latent class means $\mu_{1,i}$ and $\mu_{2,i}$ can be modeled following Eq.(2.11) as

$$\begin{aligned}
\log(\mu_{1,i}) &= \boldsymbol{x}_i'\boldsymbol{\alpha}, \\
\log(\mu_{2,i}) &= \boldsymbol{x}_i'\boldsymbol{\gamma}, \\
\text{logit}(\pi^*) &= \tau,
\end{aligned} \tag{2.17}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are $p \times 1$ vector of and parameters corresponding to component-1 and component-2, respectively, and τ is the logit transformation of the mixing proportion ($-\infty < \tau < \infty$). Note that in both latent components, same set of covariates, $\boldsymbol{x}_i$ for the ith subject is used for simplicity. The parameters of models in Eq.(2.17) can be estimated by applying EM algorithm (Dempster et al., 1977; Wang et al., 1996; Benecha et al., 2017; McLachlan and Peel, 2000).

2.3 Marginalized Poisson-Poisson Mixture Model

The interpretations of parameters in terms of incidence rate ratio (IRR) under Poisson-Poisson mixture model are often imprecise or misleading because of their latent class formulation. To overcome this problem, it is preferable to draw conclusions from the marginal mean rather than the means of the two latent classes. In order to develop a model for marginal mean of Poisson-Poisson mixture distribution, Benecha et al. (2017) suggested to express component-2 mean, using Eq.(2.15) as

$$\mu_{2,i} = \frac{\mu_i - \pi^* \mu_{1,i}}{1 - \pi^*}. \tag{2.18}$$

Then using Eq.(2.18) the variance of Poisson-Poisson mixture distribution, given in Eq.(2.16), can be expressed as

$$\text{Var}(Y_i) = \mu_i + \left[\frac{\pi^*}{1 - \pi^*}\right](\mu_i - \mu_{1,i})^2. \tag{2.19}$$

It can be observed that the marginal mean μ_i defined in Eq.(2.15) under regression models for latent class means $\mu_{1,i}$ and $\mu_{2,i}$ given in Eq.(2.17) generally depends upon a complicated function of the regression parameters . Consequently, the inference regarding marginal means is hardly possible from the FMM setup. Therefore, a new marginalized model is required to draw inference regarding the effects of covariates on μ_i directly. As suggested by Benecha et al. (2017), the marginalized Poisson-Poisson (MPois-Pois) distribution for the random variable Y_i can be obtained from Eq.(2.13) using Eq.(2.18). Therefore, the pmf of MPois-Pois random variable becomes

$$f(Y_i = y_i | \pi^*, \mu_i, \mu_{1,i}) = \pi^* \frac{e^{-\mu_{1,i}} {\mu_{1,i}}^{y_i}}{y_i!} + (1 - \pi^*) \frac{e^{-\frac{\mu_i - \pi \mu_{1,i}}{1 - \pi^*}} \left(\frac{\mu_i - \pi^* \mu_{1,i}}{1 - \pi^*}\right)^{y_i}}{y_i!}. \tag{2.20}$$

Under regression setup, the models for Eq.(2.20) can be written as (Benecha et al., 2017)

$$\begin{aligned} \log(\mu_i) &= \boldsymbol{x}_i'\boldsymbol{\beta}, \\ \log(\mu_{1,i}) &= \boldsymbol{x}_i'\boldsymbol{\alpha}, \\ \text{logit}(\pi^*) &= \tau, \end{aligned} \tag{2.21}$$

where $\boldsymbol{x}_i$ is the set of covariates of order $p \times 1$ used for ith subject in the log linear models of Eq.(2.21); also $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ both are the $p \times 1$ vectors of parameters corresponding to marginal mean and component-1 mean, respectively.

The IRR for the kth covariate in a MPois-Pois model is the ratio of means for a one-unit increase in x_{ik}, which is obtained as follows

$$\text{IRR}^k = \frac{\mu_i(x_{ik} = a + 1, \tilde{\boldsymbol{x}}_{\boldsymbol{i}})}{\mu_i(x_{ik} = a, \tilde{\boldsymbol{x}}_{\boldsymbol{i}})} = \exp(\beta_k), \tag{2.22}$$

where $\tilde{\boldsymbol{x}}_{\boldsymbol{i}}$ indicates all covariates except x_{ik}. Although, the primary interest is to estimate the parameters ($\boldsymbol{\beta}$) of the marginal mean (μ_i) of regression model as in Eq.(2.21), the estimate of nuisance parameters $\boldsymbol{\alpha}$ and τ are required to facilitate the maximum likelihood estimation of $\boldsymbol{\beta}$.

2.3.1 Likelihood Function

In order to estimate the parameters $\tau, \boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ of MPois-Pois model under maximum likelihood framework, it is required to construct the likelihood function. Let us consider $\boldsymbol{\theta} = (\tau, \boldsymbol{\beta}', \boldsymbol{\alpha}')'$ for simplicity. Suppose that there are n observations $y_1, \cdots, y_n$ which are realization of the corresponding n independent and identically distributed (iid) random variables $Y_1, \cdots, Y_n$ having probability mass function (pmf) as in Eq.(2.20). Then the likelihood function of MPois-Pois model is as follows

$$
\begin{aligned}
L(\boldsymbol{\theta}|\boldsymbol{y}) &= \prod_{i=1}^{n} f(Y_i = y_i|\boldsymbol{\theta}) \\
&= \prod_{i=1}^{n} \Big[A_i + B_i \times C_i \Big],
\end{aligned} \tag{2.23}
$$

where

$$
\begin{aligned}
A_i =& \left(\frac{1}{1+e^{-\tau}}\right)\frac{e^{-\exp(\boldsymbol{x_i'\alpha})}\left(e^{\boldsymbol{x_i'\alpha}}\right)^{y_i}}{y_i!}, \\
B_i =& \left(\frac{1}{1+e^{\tau}}\right)\frac{\exp\Bigg(-(1+e^{\tau})\Big(e^{\boldsymbol{x_i'\beta}} - \big(\frac{1}{1+e^{-\tau}}\big)e^{\boldsymbol{x_i'\alpha}}\Big)\Bigg)}{y_i!}, \\
C_i =& \Bigg((1+e^{\tau})\Big(e^{\boldsymbol{x_i'\beta}} - \big(\frac{1}{1+e^{-\tau}}\big)e^{\boldsymbol{x_i'\alpha}}\Big)\Bigg)^{y_i}.
\end{aligned}
$$

The likelihood function can also be expressed as

$$
L(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{n} \frac{1}{(1+e^{\tau})y_i!}\Bigg\{ e^{\tau}\exp(-e^{\boldsymbol{x_i'\alpha}})e^{\boldsymbol{x_i'\alpha} y_i} + e^{-\eta_i(\boldsymbol{\theta})}\eta_i(\boldsymbol{\theta})^{y_i} \Bigg\}, \tag{2.24}
$$

where $\eta_i(\boldsymbol{\theta}) = e^{\boldsymbol{x_i'\beta}}(1+e^{\tau}) - e^{\tau}e^{\boldsymbol{x_i'\alpha}}$. Therefore, the log-likelihood function of MPois-Pois model can be obtained as

$$
l(\boldsymbol{\theta}) = \sum_{i=1}^{n}\Bigg[-\log(1+e^{\tau}) - \log(y_i!) + \log\Big\{ e^{\tau - e^{\boldsymbol{x_i'\alpha}} + \boldsymbol{x_i'\alpha} y_i} + e^{-\eta_i(\boldsymbol{\theta})}\eta_i(\boldsymbol{\theta})^{y_i} \Big\} \Bigg]. \tag{2.25}
$$

2.3.2 Score Function

The score function of MPois-Pois model can be defined as

$$
U(\boldsymbol{\theta}) = \left[\frac{\partial l(\boldsymbol{\theta})}{\partial \tau}, \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}'}, \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}'}\right]' = \Big[U_1, U_2', U_3'\Big]'.
$$

In order to find the elements of the score function, it would be easier if the following derivatives could be computed first,

$$
\frac{\partial \eta_i(\boldsymbol{\theta})}{\partial \tau} = e^{\tau}\big(e^{\boldsymbol{x_i'\beta}} - e^{\boldsymbol{x_i'\alpha}}\big); \frac{\partial \eta_i(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} = -e^{\boldsymbol{x_i'\alpha}+\tau}\boldsymbol{x_i}; \frac{\partial \eta_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = (1+e^{\tau})e^{\boldsymbol{x_i'\beta}}\boldsymbol{x_i}.
$$

Using these, we can compute the required derivatives for score function as

$$U_1 = \sum_{i=1}^{n} \left[1 - \frac{e^\tau}{1+e^\tau} + \frac{e^\tau \left(e^{\boldsymbol{x}_i'\boldsymbol{\beta}} - e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}\right)\left[y_i - \eta_i(\boldsymbol{\theta})\right] e^{-\eta_i(\boldsymbol{\theta})}\eta_i(\boldsymbol{\theta})^{y_i-1} - e^{-\eta_i(\boldsymbol{\theta})}\eta_i(\boldsymbol{\theta})^{y_i}}{e^{\tau - e^{\boldsymbol{x}_i'\boldsymbol{\alpha}} + \boldsymbol{x}_i'\boldsymbol{\alpha} y_i} + e^{-\eta_i(\boldsymbol{\theta})}\eta_i(\boldsymbol{\theta})^{y_i}} \right],$$

$$U_2 = \sum_{i=1}^{n} \left[\frac{\left(y_i - e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}\right) e^{\tau - e^{\boldsymbol{x}_i'\boldsymbol{\alpha}} + \boldsymbol{x}_i'\boldsymbol{\alpha} y_i} - e^{-\eta_i(\boldsymbol{\theta})}\eta_i(\boldsymbol{\theta})^{y_i-1}\left[y_i - \eta_i(\boldsymbol{\theta})\right] e^{\boldsymbol{x}_i'\boldsymbol{\alpha}+\tau}}{e^{\tau - e^{\boldsymbol{x}_i'\boldsymbol{\alpha}} + \boldsymbol{x}_i'\boldsymbol{\alpha} y_i} + e^{-\eta_i(\boldsymbol{\theta})}\eta_i(\boldsymbol{\theta})^{y_i}} \right] \boldsymbol{x_i},$$

$$U_3 = \sum_{i=1}^{n} \left[\frac{e^{-\eta_i(\boldsymbol{\theta})}\eta_i(\boldsymbol{\theta})^{y_i-1}\left[y_i - \eta_i(\boldsymbol{\theta})\right](1+e^\tau) e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{e^{\tau - e^{\boldsymbol{x}_i'\boldsymbol{\alpha}} + \boldsymbol{x}_i'\boldsymbol{\alpha} y_i} + e^{-\eta_i(\boldsymbol{\theta})}\eta_i(\boldsymbol{\theta})^{y_i}} \right] \boldsymbol{x_i}.$$

Then the maximum likelihood estimating equations for $\boldsymbol{\theta}$ can be formed as

$$U(\boldsymbol{\theta}) = \mathbf{0}_{(2p+1)\times 1}. \tag{2.26}$$

The solution of Eq.(2.26) can be obtained by Newton's method. At the rth step, Newton's method updates the values of the parameters as

$$\boldsymbol{\theta}^{(r)} = \boldsymbol{\theta}^{(r-1)} + \left(I(\boldsymbol{\theta}^{(r-1)})\right)^{-1} U(\boldsymbol{\theta}^{(r-1)}), \tag{2.27}$$

for $r = 1, 2, \cdots$ until convergence. In Eq.(2.27), $I(\cdot)$ is a $(2p+1) \times (2p+1)$ matrix of observed information obtained from negative of Hessian. The Hessian, $H(\cdot)$, is a matrix of second partial derivative of the objective function $\ell(\boldsymbol{\theta})$, which is described in Appendix A.1.

With carefully chosen starting values, the MLE of the parameters ($\hat{\boldsymbol{\theta}}$) are the solution of Eq.(2.26) and can be obtained by the use of quasi-Newton optimization method or the Newton's method. The quasi-Newton optimization can be implemented by SAS 'nlmixed' or R 'optim' function. Starting values for τ and $\boldsymbol{\alpha}$ can be obtained by fitting the Poisson-Poisson mixture model (Benecha et al., 2017) from Eq.(2.17) by applying the EM algorithm (Dempster et al., 1977; Leisch, 2004). The EM algorithm for estimating parameters of Poisson-Poisson mixture regression model of Eq.(2.17) is given in Appendix A.2. Also, the initial values of $\boldsymbol{\beta}$ are the fitted values of the marginal parameters for MZIP model from Eq.(2.7) or the fitted values of standard Poisson regression model.

Let $I^{ll}(\boldsymbol{\theta}), l = 1, 2, \ldots, (2p+1)$ be the (l, l)th element of the inverse of information matrix. Then the standard error of the lth component of the estimator of the parameter vector $\boldsymbol{\theta}$ is given by

$$\mathrm{se}(\hat{\boldsymbol{\theta}}_l) = \sqrt{I^{ll}(\boldsymbol{\theta})}. \tag{2.28}$$

For lth parameter, $l = 1, 2, \ldots, (2p+1)$, the asymptotic behavior of the estimator can be expressed as

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_l - \boldsymbol{\theta}_l\big) \sim N\big(0, \mathrm{se}(\hat{\boldsymbol{\theta}}_l)^2\big).$$

2.4 Simulation Study

Extensive simulation studies have been carried out with varying sample sizes and mixing probabilities to investigate the fitting performance of MPois-Pois model. Other count models such as MZIP, ZIP, PH, Poisson and negative binomial models along with the MPois-Pois model have also been fitted in order to make comparison among the models in cross-sectional setup. The simulation studies were conducted in the regression setup for MPois-Pois data generating process (dgp). Note that, the data were generated using a two components Poisson mixture model using varying mixing probabilities such as $\pi^* = 0.50, 0.70, 0.90$ and we were interested in the marginal means. It is assumed that both the Poisson models were influenced by the same set of known covariates $\boldsymbol{x}_i = (x_{0i}, x_{1i}, x_{2i}, ..., x_{(p-1)i})'$ with $x_{0i} = 1$ for the ith response. In order to generate zero-inflated count data, the parameters were chosen in such a way that $0 < \mu_{1,i} < 0.50$ and large value of $\mu_{2,i}$ can be observed. To obtain marginal inference from Poisson-Poisson mixture distribution, the zero-inflated data for a sample of size n have been generated by using the following steps.

1. The covariates (x_{1i}, x_{2i}) were generated from unif(0,1), Bernoulli(0.40), respectively.

2. Suitable values of regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ were used in Eq.(2.21) to compute the marginal mean $\mu_i, i = 1, ..., n$ for zero-inflated data. Note that $\boldsymbol{\beta} = (0.20, 0.80, 0.80)'$, $\boldsymbol{\beta} = (0.20, 0.70, 0.50)'$, $\boldsymbol{\beta} = (0.15, 0.15, 0.15)'$ were used for $\pi^* = 0.50, 0.70, 0.90$, respectively.

3. Binary observations $d_1, ..., d_n$ were generated using $D \sim$Bernoulli(π^*).

4. If $d_i = 1$, suitable values of regression parameters $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)'$ were used in Eq.(2.21) to compute the component-1 means $(\mu_{1,i})$ for generating the observations $y_i \sim$ Pois$(\mu_{1,i})$. Note that $\boldsymbol{\alpha} = (-1.00, 0.20, 0.10)'$ were used for each of $\pi^* = 0.50, 0.70, 0.90$. If $\mu_{1,i} \geq 0.50$ was observed in this process, a set of values $\boldsymbol{x}_l = (x_{0l}, x_{1l}, x_{2l})'$ for which $\mu_{1,l} < 0.50$ had been observed was replaced in place of the set of values $\boldsymbol{x}_i = (x_{0i}, x_{1i}, x_{2i})'$.

5. If $d_i = 0$, the relationship $\mu_{2,i} = \dfrac{\mu_i - \pi\mu_{1,i}}{1 - \pi}$ from Eq.(2.18) was used to compute the component-2 means $(\mu_{2,i})$ for generating $y_i \sim$ Pois$(\mu_{2,i})$.

Using the generated data, the regression parameters of MPois-Pois model were estimated by employing maximum likelihood (ML) approach. Other count models viz. MZIP, negative binomial (NB), Poisson (Pois), ZIP, PH models have also been fitted to the generated data by using the same (ML) approach.

The simulation was repeated 2000 times for each setup. In order to investigate the performance of ML estimates, we have computed the biases, standard errors and the coverage probability (Cov.Pr.) by using these repetitions. The biases were computed from the differences between simulated means (SM) and the true values for each of the parameters. Two types of standard errors such as estimated standard errors (ESE) and simulated standard errors (SSE) were also computed to investigate the properties of the estimators. The SM, ESE and SSE for a estimator $\hat{\theta}$ were defined respectively as

$$\hat{\theta} = \frac{1}{r}\sum_{l=1}^{r}\hat{\theta}_l; ESE_{\hat{\theta}} = \frac{1}{r}\sum_{l=1}^{r} se_l(\hat{\theta}); \quad \text{and} \quad SSE_{\hat{\theta}} = \sqrt{\frac{1}{r-1}\sum_{l=1}^{r}(\hat{\theta}_l - \hat{\hat{\theta}})^2}, \tag{2.29}$$

where r represents the number of repetitions in the simulation and $se(\hat{\theta})$ was computed using Eq. (2.28). The proportion of convergences (Conv.Prop.) in fitting the MPois-Pois model were computed for all the setups. The results obtained from the fitted MPois-Pois model for π^* =0.90, 0.70 and 0.50 are shown in Table 2.1. Also, the MZIP, ZIP, PH, NB, and Poisson models have been fitted for making comparison with MPois-Pois model. The results obtained from these models for π^* =0.90, and 0.50 are shown in Table 2.2.

From Table 2.1, it is clear that the estimates of marginal parameters had minimal amount of biases for all the settings except for sample size 100 with $\pi^* = 0.90$. These biases decrease with increasing the sample size. For example, the amount of biases of $(\beta_0, \beta_1, \beta_2)$ were (-0.023,0.026,-0.308), (-0.027,0.020,-0.035), (-0.018,0.004,-0.001) for $\pi^* = 0.90, 0.70, 0.50$, respectively when $n = 100$; (-0.011,0.004,0.005), (-0.009,0.006,0.001), (-0.001,-0.004,-0.003) for $\pi^* = 0.90, 0.70, 0.50$, respectively when $n = 500$; and (-0.003,0.000,-0.001), (-0.002,0.002,0.000), (0.001,-0.002,-0.002) for $\pi^* = 0.90, 0.70, 0.50$, respectively when $n = 2000$.

The largest standard errors of all the estimated parameters were found for $\pi^* = 0.90$ when sample size 100. The standard errors decrease with increasing the sample size. Also, for a given sample size, the standard errors of all the estimated parameters for $\pi^* = 0.50$

are smaller than for $\pi^* = 0.90$. The smallest standard errors of all the estimated parameters were found for $\pi^* = 0.50$ when sample size 2000.

The coverage probabilities of the confidence interval constructed from the estimates of the marginal models were found to be approximately equal to the nominal level of confidence in all the setups.

Table 2.1: Simulated mean (SM), simulated mean of mixing proportion (SMP), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr.) for estimating Marginal Parameters ($\boldsymbol{\beta}$); Component-1 Parameters ($\boldsymbol{\alpha}$) and τ under MPois-Pois model for MPois-Pois dgp

π^*	n	Conv.Prop.	Params	SM	SMP	Bias	ESE	SSE	Cov.Pr.
0.90	100	0.949	$\beta_0 = 0.15$	0.127		-0.023	0.318	0.367	93.2
			$\beta_1 = 0.15$	0.176		0.026	0.393	0.493	92.9
			$\beta_2 = 0.15$	-0.158		-0.308	0.374	0.683	79.1
			$\alpha_0 = -1.0$	-1.087	0.897	-0.087	0.397	0.411	95.7
			$\alpha_1 = 0.20$	0.256		0.056	0.621	0.650	94.7
			$\alpha_2 = 0.10$	0.032		-0.068	0.628	0.603	97.6
			$\tau = 2.1972$	2.162		-0.035	0.353	0.339	94.9
	500	1.000	$\beta_0 = 0.15$	0.139		-0.011	0.133	0.136	94.7
			$\beta_1 = 0.15$	0.154		0.004	0.174	0.174	95.2
			$\beta_2 = 0.15$	0.155		0.005	0.117	0.117	95.6
			$\alpha_0 = -1.0$	-1.014	0.901	-0.014	0.171	0.168	95.5
			$\alpha_1 = 0.20$	0.201		0.001	0.300	0.296	95.4
			$\alpha_2 = 0.10$	0.096		-0.004	0.208	0.209	95.5
			$\tau = 2.1972$	2.204		0.007	0.153	0.156	94.9
	1000	1.000	$\beta_0 = 0.15$	0.146		-0.004	0.096	0.097	94.6
			$\beta_1 = 0.15$	0.154		0.004	0.123	0.124	95.4
			$\beta_2 = 0.15$	0.151		0.001	0.077	0.076	95.0
			$\alpha_0 = -1.0$	-1.009	0.900	-0.009	0.126	0.128	95.1
			$\alpha_1 = 0.20$	0.206		0.006	0.215	0.220	94.6
			$\alpha_2 = 0.10$	0.098		-0.002	0.138	0.137	95.3
			$\tau = 2.1972$	2.199		0.002	0.107	0.108	95.3
	2000	1.000	$\beta_0 = 0.15$	0.147		-0.003	0.067	0.069	95.0
			$\beta_1 = 0.15$	0.150		0.000	0.085	0.086	94.5
			$\beta_2 = 0.15$	0.149		-0.001	0.058	0.058	95.2
			$\alpha_0 = -1.0$	-1.003	0.900	-0.003	0.089	0.090	94.5
			$\alpha_1 = 0.20$	0.199		-0.001	0.148	0.148	95.4
			$\alpha_2 = 0.10$	0.098		-0.002	0.104	0.105	93.9
			$\tau = 2.1972$	2.200		0.003	0.076	0.076	94.4
0.70	100	0.948	$\beta_0 = 0.20$	0.173		-0.027	0.250	0.262	94.1
			$\beta_1 = 0.70$	0.720		0.020	0.368	0.397	93.5
			$\beta_2 = 0.50$	0.465		-0.035	0.307	0.380	92.5
			$\alpha_0 = -1.0$	-1.099	0.699	-0.099	0.609	0.701	94.5
			$\alpha_1 = 0.20$	0.222		0.022	0.895	1.003	94.7
			$\alpha_2 = 0.10$	0.048		-0.052	0.825	0.805	97.3
			$\tau = 0.8473$	0.841		-0.006	0.266	0.265	95.4
	500	1.000	$\beta_0 = 0.20$	0.191		-0.009	0.110	0.114	94.0
			$\beta_1 = 0.70$	0.706		0.006	0.177	0.181	94.3
			$\beta_2 = 0.50$	0.501		0.001	0.111	0.110	96.0
			$\alpha_0 = -1.0$	-1.027	0.700	-0.027	0.250	0.247	95.7
			$\alpha_1 = 0.20$	0.219		0.019	0.408	0.399	95.6
			$\alpha_2 = 0.10$	0.097		-0.003	0.274	0.271	96.0
			$\tau = 0.8473$	0.847		0.000	0.117	0.116	94.9
	1000	1.000	$\beta_0 = 0.20$	0.196		-0.004	0.080	0.079	94.8
			$\beta_1 = 0.70$	0.701		0.001	0.124	0.123	94.8

Continued...Table 2.1

π^*	n	Conv.Prop.	Params	SM	SMP	Bias	ESE	SSE	Cov.Pr.
			$\beta_2 = 0.50$	0.500		0.000	0.075	0.074	95.5
			$\alpha_0 = -1.0$	-1.012	0.700	-0.012	0.181	0.183	95.0
			$\alpha_1 = 0.20$	0.205		0.005	0.289	0.294	94.9
			$\alpha_2 = 0.10$	0.098		-0.002	0.182	0.181	94.9
			$\tau = 0.8473$	0.848		0.000	0.081	0.082	95.0
	2000	1.000	$\beta_0 = 0.20$	0.198		-0.002	0.057	0.056	94.6
			$\beta_1 = 0.70$	0.702		0.002	0.087	0.087	94.9
			$\beta_2 = 0.50$	0.500		-0.000	0.057	0.058	94.6
			$\alpha_0 = -1.0$	-1.009	0.700	-0.009	0.128	0.131	94.3
			$\alpha_1 = 0.20$	0.210		0.010	0.200	0.208	94.0
			$\alpha_2 = 0.10$	0.100		0.000	0.137	0.137	95.9
			$\tau = 0.8473$	0.848		0.001	0.057	0.060	94.3
0.50	100	0.850	$\beta_0 = 0.20$	0.182		-0.018	0.225	0.243	92.7
			$\beta_1 = 0.80$	0.804		0.004	0.346	0.368	93.2
			$\beta_2 = 0.80$	0.799		-0.001	0.269	0.305	92.8
			$\alpha_0 = -1.0$	-1.655	0.505	-0.655	5.493	5.494	91.4
			$\alpha_1 = 0.20$	0.649		0.449	1.914	5.910	94.1
			$\alpha_2 = 0.10$	0.532		0.432	5.527	3.699	95.6
			$\tau = 0.0$	0.019		0.019	0.323	0.325	93.9
	500	1.000	$\beta_0 = 0.20$	0.199		-0.001	0.100	0.101	94.8
			$\beta_1 = 0.80$	0.796		-0.004	0.169	0.170	94.4
			$\beta_2 = 0.80$	0.797		-0.003	0.105	0.104	95.5
			$\alpha_0 = -1.0$	-1.027	0.500	-0.027	0.375	0.396	94.6
			$\alpha_1 = 0.20$	0.196		-0.004	0.590	0.608	95.2
			$\alpha_2 = 0.10$	0.081		-0.019	0.382	0.377	96.4
			$\tau = 0.0$	0.000		0.000	0.141	0.143	94.7
	1000	1.000	$\beta_0 = 0.20$	0.199		-0.001	0.073	0.074	94.8
			$\beta_1 = 0.80$	0.799		-0.001	0.119	0.120	95.0
			$\beta_2 = 0.80$	0.799		-0.001	0.071	0.073	93.8
			$\alpha_0 = -1.0$	-1.009	0.500	-0.009	0.265	0.271	94.8
			$\alpha_1 = 0.20$	0.190		-0.010	0.409	0.414	94.9
			$\alpha_2 = 0.10$	0.094		-0.006	0.249	0.258	94.5
			$\tau = 0.0$	-0.002		-0.002	0.095	0.094	95.3
	2000	1.000	$\beta_0 = 0.20$	0.201		0.001	0.052	0.052	94.8
			$\beta_1 = 0.80$	0.798		-0.002	0.084	0.083	95.2
			$\beta_2 = 0.80$	0.798		-0.002	0.054	0.054	95.3
			$\alpha_0 = -1.0$	-1.005	0.500	-0.005	0.185	0.186	95.1
			$\alpha_1 = 0.20$	0.197		-0.003	0.281	0.286	94.5
			$\alpha_2 = 0.10$	0.100		0.000	0.186	0.190	95.3
			$\tau = 0.0$	0.002		0.002	0.068	0.069	95.1

Table 2.2: Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr.) for estimating Marginal Parameters: ($\boldsymbol{\beta}$) under MZIP, NB, Pois, ZIP and PH models along with Component-1 Parameters ($\boldsymbol{\alpha}$) for MZIP model for MPois-Pois dgp

π^*	Model	n	Params	SM	Bias	ESE	SSE	Cov.Pr.
0.90	MZIP	100	$\beta_0 = 0.15$	0.082	-0.068	0.329	0.509	80.5
			$\beta_1 = 0.15$	0.173	0.023	0.519	0.797	80.2
			$\beta_2 = 0.15$	-0.014	-0.164	0.478	0.843	73.3
			$\alpha_0 = -1.0$	0.269	1.269	0.705	2.397	20.1
			$\alpha_1 = 0.20$	-0.158	-0.358	0.867	2.483	89.3
			$\alpha_2 = 0.10$	-1.673	-1.773	32.581	3.927	96.6
		500	$\beta_0 = 0.15$	0.132	-0.018	0.147	0.218	80.5
			$\beta_1 = 0.15$	0.156	0.006	0.258	0.378	82.3
			$\beta_2 = 0.15$	0.144	-0.006	0.179	0.279	79.5
			$\alpha_0 = -1.0$	0.405	1.405	0.211	0.221	0.0
			$\alpha_1 = 0.20$	-0.198	-0.398	0.379	0.395	80.1
			$\alpha_2 = 0.10$	-0.091	-0.191	0.265	0.281	88.2
		1000	$\beta_0 = 0.15$	0.140	-0.010	0.109	0.166	80.3
			$\beta_1 = 0.15$	0.156	0.006	0.186	0.282	79.4
			$\beta_2 = 0.15$	0.152	0.002	0.119	0.183	79.9
			$\alpha_0 = -1.0$	0.414	1.414	0.156	0.164	0.0
			$\alpha_1 = 0.20$	-0.210	-0.410	0.271	0.289	66.3
			$\alpha_2 = 0.10$	-0.084	-0.184	0.174	0.179	80.5
		2000	$\beta_0 = 0.15$	0.145	-0.005	0.076	0.114	80.3
			$\beta_1 = 0.15$	0.150	0.000	0.128	0.194	79.8
			$\beta_2 = 0.15$	0.148	-0.002	0.090	0.143	79.6
			$\alpha_0 = -1.0$	0.410	1.410	0.109	0.114	0.0
			$\alpha_1 = 0.20$	-0.197	-0.397	0.186	0.192	43.4
			$\alpha_2 = 0.10$	-0.081	-0.181	0.130	0.135	72.1
	NB	100	$\beta_0 = 0.15$	0.083	-0.067	0.440	0.521	91.8
			$\beta_1 = 0.15$	0.172	0.022	0.709	0.820	90.7
			$\beta_2 = 0.15$	-0.013	-0.163	0.668	0.848	88.3
		500	$\beta_0 = 0.15$	0.131	-0.019	0.203	0.219	93.2
			$\beta_1 = 0.15$	0.157	0.007	0.364	0.381	93.8
			$\beta_2 = 0.15$	0.144	-0.006	0.254	0.279	92.9
		1000	$\beta_0 = 0.15$	0.140	-0.010	0.151	0.166	92.6
			$\beta_1 = 0.15$	0.156	0.006	0.263	0.282	93.5
			$\beta_2 = 0.15$	0.152	0.002	0.169	0.183	92.5
		2000	$\beta_0 = 0.15$	0.145	-0.005	0.106	0.114	93.8
			$\beta_1 = 0.15$	0.151	0.001	0.182	0.195	93.1
			$\beta_2 = 0.15$	0.148	-0.002	0.127	0.143	91.1
	Pois	100	$\beta_0 = 0.15$	0.084	-0.066	0.202	0.507	57.9
			$\beta_1 = 0.15$	0.173	0.023	0.321	0.791	58.8
			$\beta_2 = 0.15$	-0.016	-0.166	0.322	0.843	51.6
		500	$\beta_0 = 0.15$	0.132	-0.018	0.089	0.218	58.1
			$\beta_1 = 0.15$	0.156	0.006	0.157	0.378	58.6

Continued...Table 2.2

π^*	Model	n	Params	SM	Bias	ESE	SSE	Cov.Pr.
			$\beta_2 = 0.15$	0.144	-0.006	0.108	0.278	57.5
		1000	$\beta_0 = 0.15$	0.140	-0.010	0.066	0.166	56.2
			$\beta_1 = 0.15$	0.156	0.006	0.113	0.281	57.5
			$\beta_2 = 0.15$	0.152	0.002	0.071	0.183	56.3
		2000	$\beta_0 = 0.15$	0.145	-0.005	0.046	0.114	56.3
			$\beta_1 = 0.15$	0.151	0.001	0.078	0.194	56.9
			$\beta_2 = 0.15$	0.148	-0.002	0.054	0.143	53.5
	ZIP	100	$\beta_0 = 0.15$	0.996	0.846	0.233	0.570	20.8
			$\beta_1 = 0.15$	0.042	-0.108	0.369	0.882	57.8
			$\beta_2 = 0.15$	-0.235	-0.385	0.385	1.086	55.0
			$\alpha_0 = -1.0$	0.360	1.360	0.515	0.586	20.8
			$\alpha_1 = 0.20$	-0.221	-0.421	0.830	0.960	91.3
			$\alpha_2 = 0.10$	-1.572	-1.672	16.539	3.273	97.7
		500	$\beta_0 = 0.15$	1.050	0.900	0.098	0.225	0.7
			$\beta_1 = 0.15$	0.042	-0.108	0.173	0.392	60.5
			$\beta_2 = 0.15$	0.095	-0.055	0.118	0.289	58.3
			$\alpha_0 = -1.0$	0.405	1.405	0.211	0.219	0.0
			$\alpha_1 = 0.20$	-0.198	-0.398	0.378	0.392	80.4
			$\alpha_2 = 0.10$	-0.091	-0.191	0.265	0.280	88.3
		1000	$\beta_0 = 0.15$	1.062	0.912	0.072	0.170	0.0
			$\beta_1 = 0.15$	0.036	-0.114	0.123	0.287	57.0
			$\beta_2 = 0.15$	0.105	-0.045	0.077	0.183	58.5
			$\alpha_0 = -1.0$	0.414	1.414	0.156	0.162	0.0
			$\alpha_1 = 0.20$	-0.209	-0.409	0.271	0.286	66.5
			$\alpha_2 = 0.10$	-0.083	-0.183	0.173	0.177	80.7
		2000	$\beta_0 = 0.15$	1.064	0.914	0.051	0.113	0.0
			$\beta_1 = 0.15$	0.037	-0.113	0.085	0.193	54.0
			$\beta_2 = 0.15$	0.102	-0.048	0.058	0.142	55.8
			$\alpha_0 = -1.0$	0.410	1.410	0.109	0.114	0.0
			$\alpha_1 = 0.20$	-0.198	-0.398	0.186	0.191	43.4
			$\alpha_2 = 0.10$	-0.081	-0.181	0.130	0.135	71.9
	PH	100	$\beta_0 = 0.15$	0.994	0.844	0.235	0.581	20.5
			$\beta_1 = 0.15$	0.044	-0.106	0.372	0.895	58.2
			$\beta_2 = 0.15$	-1.114	-1.264	10.632	3.367	65.4
			$\alpha_0 = -1.0$	-0.544	0.456	0.445	0.451	80.3
			$\alpha_1 = 0.20$	0.247	0.047	0.716	0.728	95.7
			$\alpha_2 = 0.10$	0.125	0.025	0.670	0.660	97.2
		500	$\beta_0 = 0.15$	1.050	0.900	0.098	0.224	0.7
			$\beta_1 = 0.15$	0.043	-0.107	0.173	0.392	60.7
			$\beta_2 = 0.15$	0.095	-0.055	0.118	0.289	58.4
			$\alpha_0 = -1.0$	-0.510	0.490	0.200	0.197	31.0
			$\alpha_1 = 0.20$	0.202	0.002	0.358	0.351	96.1
			$\alpha_2 = 0.10$	0.102	0.002	0.251	0.252	94.9
		1000	$\beta_0 = 0.15$	1.062	0.912	0.072	0.170	0.0
			$\beta_1 = 0.15$	0.036	-0.114	0.123	0.287	57.0

Continued...Table 2.2

π^*	Model	n	Params	SM	Bias	ESE	SSE	Cov.Pr.
			$\beta_2 = 0.15$	0.105	-0.045	0.077	0.183	58.3
			$\alpha_0 = -1.0$	-0.511	0.489	0.148	0.148	9.5
			$\alpha_1 = 0.20$	0.211	0.011	0.258	0.260	94.8
			$\alpha_2 = 0.10$	0.104	0.004	0.166	0.164	94.9
		2000	$\beta_0 = 0.15$	1.064	0.914	0.051	0.113	0.0
			$\beta_1 = 0.15$	0.037	-0.113	0.085	0.193	53.9
			$\beta_2 = 0.15$	0.102	-0.048	0.058	0.142	55.8
			$\alpha_0 = -1.0$	-0.505	0.495	0.104	0.105	0.2
			$\alpha_1 = 0.20$	0.200	0.000	0.178	0.176	95.6
			$\alpha_2 = 0.10$	0.102	0.002	0.125	0.126	94.8
0.50	MZIP	100	$\beta_0 = 0.20$	0.185	-0.015	0.223	0.249	91.3
			$\beta_1 = 0.80$	0.805	0.005	0.345	0.390	90.9
			$\beta_2 = 0.80$	0.748	-0.052	0.298	0.351	91.2
			$\alpha_0 = -1.0$	-1.054	-0.054	0.711	1.000	93.9
			$\alpha_1 = 0.20$	0.235	0.035	1.024	1.288	95.0
			$\alpha_2 = 0.10$	0.009	-0.091	14.871	1.668	98.4
		500	$\beta_0 = 0.20$	0.202	0.002	0.100	0.107	93.2
			$\beta_1 = 0.80$	0.791	-0.009	0.173	0.186	92.8
			$\beta_2 = 0.80$	0.790	-0.010	0.114	0.128	91.6
			$\alpha_0 = -1.0$	-0.944	0.056	0.278	0.289	92.5
			$\alpha_1 = 0.20$	0.159	-0.041	0.458	0.462	94.6
			$\alpha_2 = 0.10$	0.160	0.060	0.295	0.291	95.3
		1000	$\beta_0 = 0.20$	0.201	0.001	0.074	0.080	92.9
			$\beta_1 = 0.80$	0.796	-0.004	0.124	0.136	92.5
			$\beta_2 = 0.80$	0.797	-0.003	0.076	0.088	91.2
			$\alpha_0 = -1.0$	-0.925	0.075	0.200	0.206	92.1
			$\alpha_1 = 0.20$	0.132	-0.068	0.323	0.327	94.6
			$\alpha_2 = 0.10$	0.144	0.044	0.197	0.205	93.5
		2000	$\beta_0 = 0.20$	0.202	0.002	0.052	0.055	92.8
			$\beta_1 = 0.80$	0.795	-0.005	0.086	0.093	92.8
			$\beta_2 = 0.80$	0.796	-0.004	0.058	0.065	91.8
			$\alpha_0 = -1.0$	-0.925	0.075	0.142	0.144	89.8
			$\alpha_1 = 0.20$	0.138	-0.062	0.225	0.227	93.1
			$\alpha_2 = 0.10$	0.152	0.052	0.149	0.151	93.0
	NB	100	$\beta_0 = 0.20$	0.181	-0.019	0.270	0.250	96.0
			$\beta_1 = 0.80$	0.812	0.012	0.420	0.391	95.8
			$\beta_2 = 0.80$	0.755	-0.045	0.364	0.349	95.6
		500	$\beta_0 = 0.20$	0.199	-0.001	0.123	0.107	97.2
			$\beta_1 = 0.80$	0.795	-0.005	0.214	0.186	97.6
			$\beta_2 = 0.80$	0.792	-0.008	0.143	0.128	96.6
		1000	$\beta_0 = 0.20$	0.198	-0.002	0.092	0.080	97.7
			$\beta_1 = 0.80$	0.801	0.001	0.156	0.136	97.4
			$\beta_2 = 0.80$	0.800	-0.000	0.097	0.088	96.9
		2000	$\beta_0 = 0.20$	0.201	0.001	0.065	0.055	97.9
			$\beta_1 = 0.80$	0.798	-0.002	0.107	0.093	97.6

Continued...Table 2.2

π^*	Model	n	Params	SM	Bias	ESE	SSE	Cov.Pr.
			$\beta_2 = 0.80$	0.797	-0.003	0.073	0.065	96.8
	Pois	100	$\beta_0 = 0.20$	0.182	-0.018	0.176	0.252	83.4
			$\beta_1 = 0.80$	0.810	0.010	0.259	0.394	80.6
			$\beta_2 = 0.80$	0.754	-0.046	0.203	0.351	76.2
		500	$\beta_0 = 0.20$	0.199	-0.001	0.078	0.109	84.3
			$\beta_1 = 0.80$	0.795	-0.005	0.129	0.189	81.5
			$\beta_2 = 0.80$	0.793	-0.007	0.078	0.130	76.1
		1000	$\beta_0 = 0.20$	0.198	-0.002	0.057	0.081	84.4
			$\beta_1 = 0.80$	0.800	0.000	0.091	0.138	81.0
			$\beta_2 = 0.80$	0.799	-0.001	0.053	0.089	76.5
		2000	$\beta_0 = 0.20$	0.200	0.000	0.041	0.056	84.3
			$\beta_1 = 0.80$	0.798	-0.002	0.064	0.094	81.4
			$\beta_2 = 0.80$	0.797	-0.003	0.040	0.066	76.5
	ZIP	100	$\beta_0 = 0.20$	0.519	0.319	0.209	0.250	58.8
			$\beta_1 = 0.80$	0.846	0.046	0.298	0.365	89.0
			$\beta_2 = 0.80$	0.798	-0.002	0.223	0.298	88.2
			$\alpha_0 = -1.0$	-1.058	-0.058	0.723	0.904	94.3
			$\alpha_1 = 0.20$	0.247	0.047	1.046	1.225	96.0
			$\alpha_2 = 0.10$	0.011	-0.089	14.476	1.610	98.9
		500	$\beta_0 = 0.20$	0.534	0.334	0.092	0.105	9.0
			$\beta_1 = 0.80$	0.837	0.037	0.145	0.172	88.4
			$\beta_2 = 0.80$	0.841	0.041	0.086	0.110	84.3
			$\alpha_0 = -1.0$	-0.946	0.054	0.281	0.291	92.6
			$\alpha_1 = 0.20$	0.162	-0.038	0.462	0.465	94.6
			$\alpha_2 = 0.10$	0.161	0.061	0.297	0.292	95.7
		1000	$\beta_0 = 0.20$	0.536	0.336	0.066	0.079	0.7
			$\beta_1 = 0.80$	0.835	0.035	0.101	0.125	87.4
			$\beta_2 = 0.80$	0.842	0.042	0.058	0.075	82.4
			$\alpha_0 = -1.0$	-0.926	0.074	0.202	0.207	92.2
			$\alpha_1 = 0.20$	0.134	-0.066	0.325	0.328	94.6
			$\alpha_2 = 0.10$	0.144	0.044	0.199	0.206	93.8
		2000	$\beta_0 = 0.20$	0.537	0.337	0.047	0.056	0.0
			$\beta_1 = 0.80$	0.836	0.036	0.072	0.088	85.9
			$\beta_2 = 0.80$	0.843	0.043	0.044	0.057	77.1
			$\alpha_0 = -1.0$	-0.927	0.073	0.143	0.145	89.8
			$\alpha_1 = 0.20$	0.141	-0.059	0.227	0.229	93.1
			$\alpha_2 = 0.10$	0.153	0.053	0.149	0.153	93.0
	PH	100	$\beta_0 = 0.20$	0.518	0.318	0.211	0.253	59.4
			$\beta_1 = 0.80$	0.847	0.047	0.300	0.370	89.4
			$\beta_2 = 0.80$	0.793	-0.007	0.297	0.397	88.5
			$\alpha_0 = -1.0$	0.400	1.400	0.447	0.466	11.0
			$\alpha_1 = 0.20$	0.397	0.197	0.733	0.756	94.1
			$\alpha_2 = 0.10$	0.418	0.318	11.332	1.462	97.6
		500	$\beta_0 = 0.20$	0.533	0.333	0.092	0.106	9.3

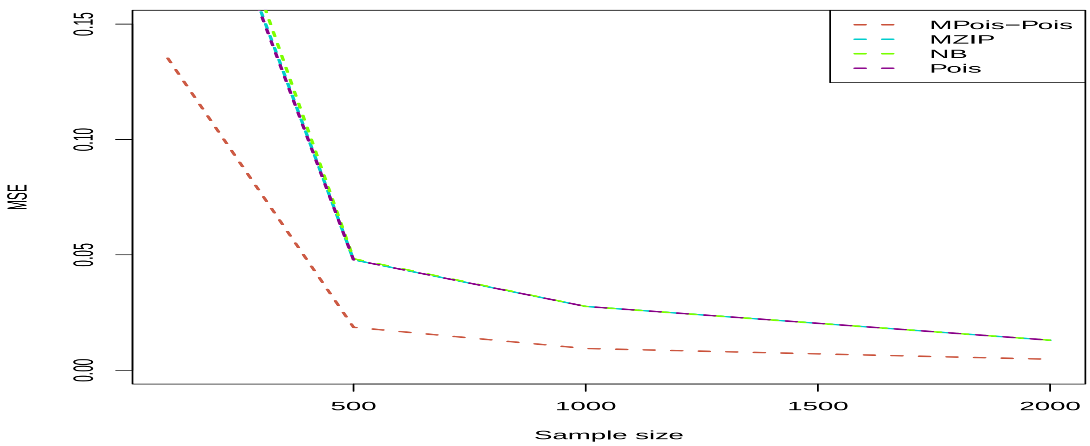
Continued...Table 2.2

π^*	Model	n	Params	SM	Bias	ESE	SSE	Cov.Pr.
			$\beta_1 = 0.80$	0.839	0.039	0.146	0.174	88.8
			$\beta_2 = 0.80$	0.842	0.042	0.086	0.110	84.2
			$\alpha_0 = -1.0$	0.410	1.410	0.201	0.205	0.0
			$\alpha_1 = 0.20$	0.365	0.165	0.366	0.363	92.4
			$\alpha_2 = 0.10$	0.224	0.124	0.260	0.254	94.2
		1000	$\beta_0 = 0.20$	0.535	0.335	0.067	0.079	0.7
			$\beta_1 = 0.80$	0.838	0.038	0.102	0.126	87.2
			$\beta_2 = 0.80$	0.843	0.043	0.058	0.075	82.4
			$\alpha_0 = -1.0$	0.414	1.414	0.149	0.152	0.0
			$\alpha_1 = 0.20$	0.356	0.156	0.265	0.265	91.4
			$\alpha_2 = 0.10$	0.227	0.127	0.172	0.179	87.7
		2000	$\beta_0 = 0.20$	0.535	0.335	0.048	0.056	0.0
			$\beta_1 = 0.80$	0.838	0.038	0.072	0.088	85.7
			$\beta_2 = 0.80$	0.844	0.044	0.045	0.057	76.4
			$\alpha_0 = -1.0$	0.410	1.410	0.105	0.104	0.0
			$\alpha_1 = 0.20$	0.361	0.161	0.182	0.181	86.4
			$\alpha_2 = 0.10$	0.232	0.132	0.129	0.132	82.6

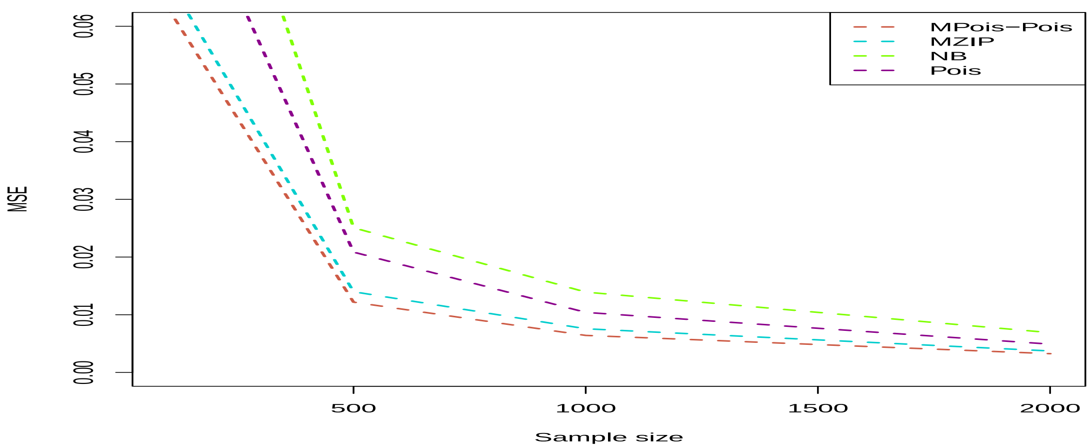
To compare the estimates of all the parameters of MPois-Pois model with the estimates of all the parameters of some other count models such as MZIP, negative binomial and Poisson models , we employed the mean square error (MSE) criterion. The sample size versus MSE were plotted in Figures 2.1-2.3 for estimates of β_0, β_1 and β_2, respectively. All the figures were plotted in three parts, part (a) is for $\pi^* = 0.90$, part (b) is for $\pi^* = 0.70$ and part (c) is for $\pi^* = 0.50$.

From Figures 2.1-2.3, it was clear that the MSE of estimators from MPois-Pois model for all the parameters were lowest compared to others for all of setups. For a given sample size and a given parameter, the MSE criterion indicates that the MPois-Pois model will perform better if the mixing proportion (π^*) increases towards 1. In other words, the MPois-Pois provides the best-fit model when a larger proportion of counts arise from the latent group of mixtures with smaller means.

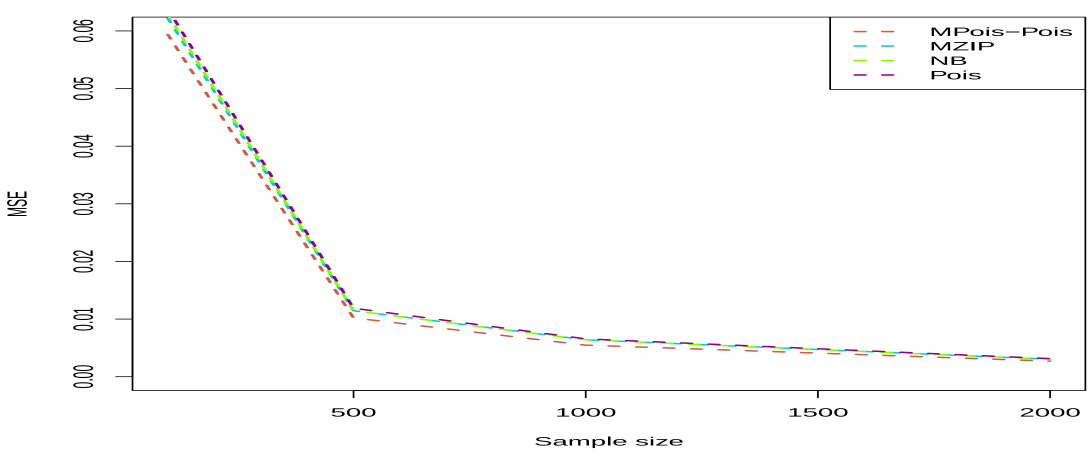
From the comparative study, it is obvious that when data arise from mixture of two-component Poisson distribution, marginal inference of the parameters from the available count models may provide larger amount of biases and/or higher standard errors of the estimates. However, the use of the MPois-Pois model in such instance provides consistent as well as efficient estimates of the marginal parameters of interest.



(a) Sample size vs. MSE for $\pi^* = 0.90$

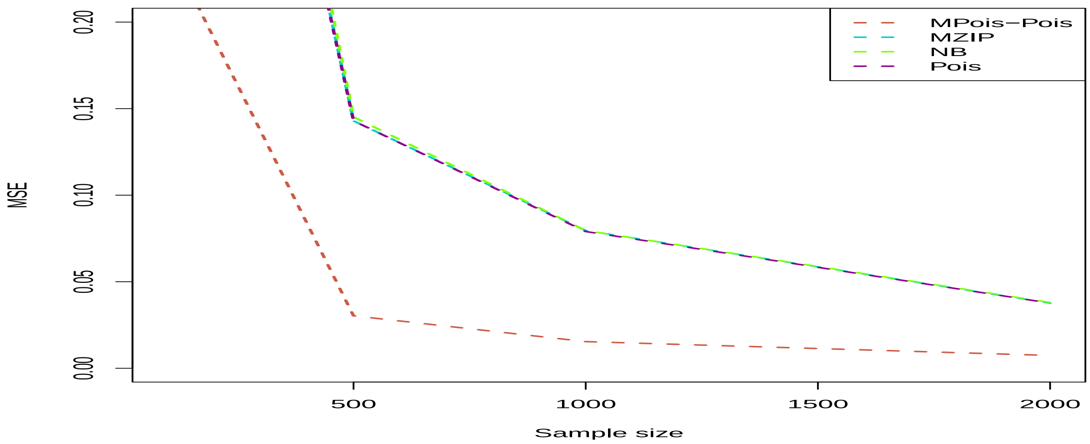


(b) Sample size vs. MSE for $\pi^* = 0.70$

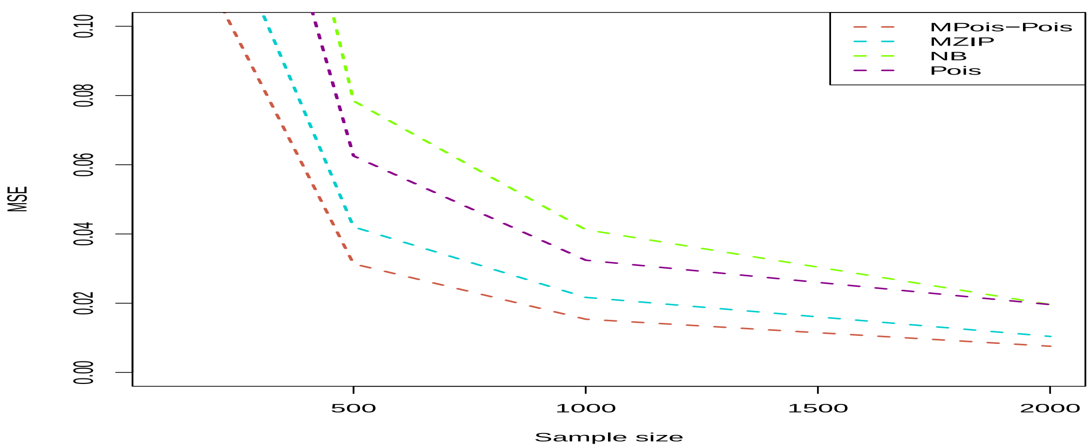


(c) Sample size vs. MSE for $\pi^* = 0.50$

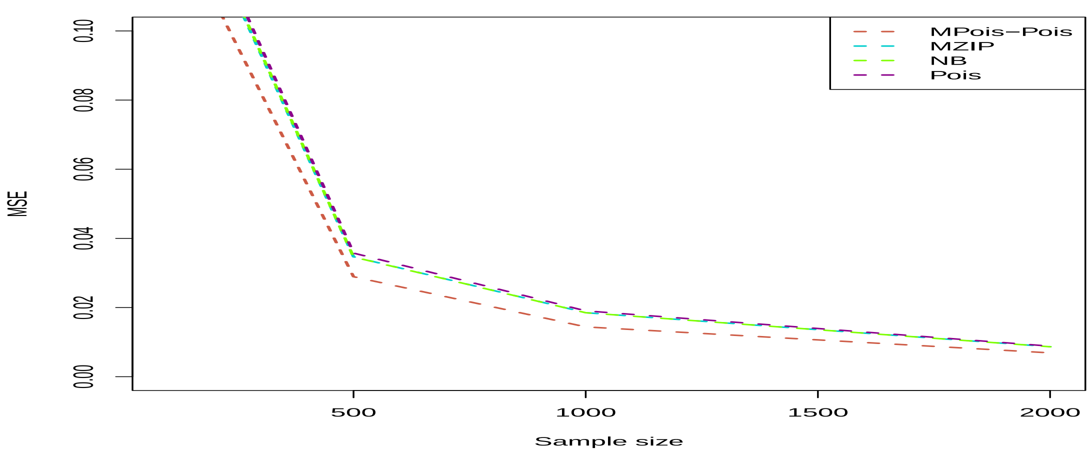
Figure 2.1: A comparison of MSE estimated from MPois-Pois, MZIP, negative binomial (NB) and Poisson (Pois) models with varying sample sizes for regression parameter β_0



(a) Sample size vs. MSE for $\pi^* = 0.90$

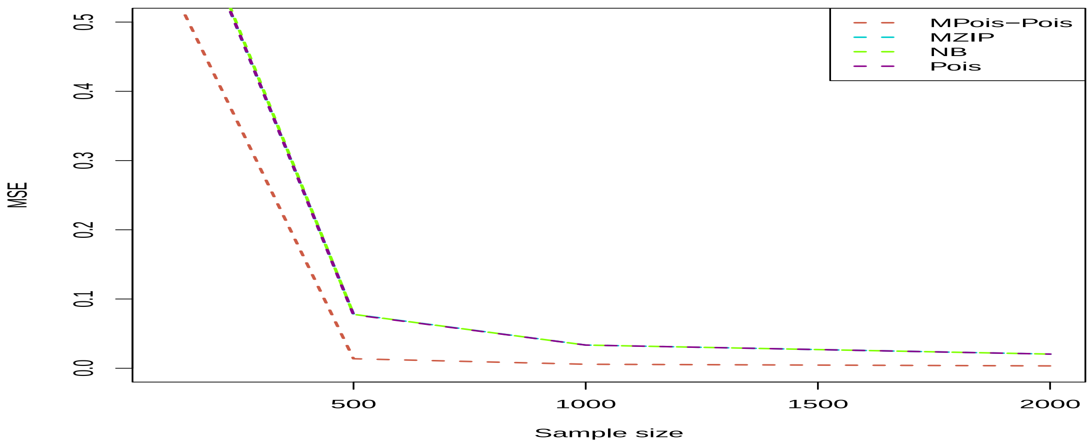


(b) Sample size vs. MSE for $\pi^* = 0.70$

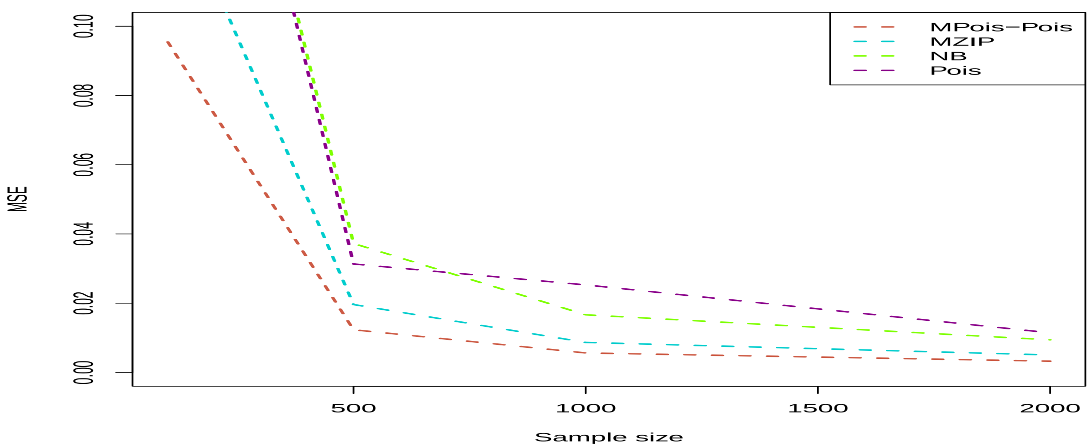


(c) Sample size vs. MSE for $\pi^* = 0.50$

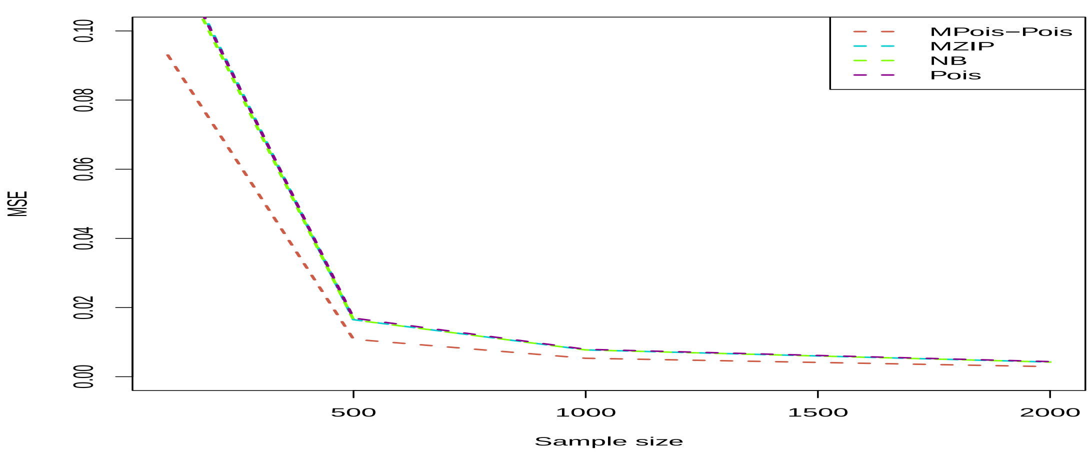
Figure 2.2: A comparison of MSE estimated from MPois-Pois, MZIP, negative binomial (NB) and Poisson (Pois) models with varying sample sizes for regression parameter β_1



(a) Sample size vs. MSE for $\pi^* = 0.90$



(b) Sample size vs. MSE for $\pi^* = 0.70$



(c) Sample size vs. MSE for $\pi^* = 0.50$

Figure 2.3: A comparison of MSE estimated from MPois-Pois, MZIP, negative binomial (NB) and Poisson (Pois) models with varying sample sizes for regression parameter β_2

2.5 Illustration

2.5.1 Data

To illustrate the application of MPois-Pois mixture model, a nationwide data extracted from the 2014 Bangladesh Demographic and Health Survey (BDHS) have been utilized. This survey consists of a two stage sampling. In the first stage of sampling, 600 clusters were randomly selected, 393 from rural and 207 from urban areas. In the second stage, about 30 households were randomly selected from each cluster using systematic sampling. Based on the design, A total of 17,863 ever married women of reproductive age were interviewed to collect data. Then the interviewers collect data on fertility, family planning along with socioeconomic and demographic characteristics. The interviewers also collect data on several aspect of maternal and newborn health, including antenatal care (ANC). Women having children born in preceding three years of the survey were only considered and information regarding ANC of the children was collected from the most recent births.

2.5.2 Variables

The outcome variable considered is *the number of ANC visits* a woman received during her most recent pregnancy period before the survey. Some important covariates were included in the analysis based on some available literature (Haque et al., 2022a; Bhowmik et al., 2020; Islam and Masud, 2018; Hossain et al., 2020). These variables are: area of residence (*urban, rural*); level of education (*no education, primary, secondary, higher*); media exposure for any of the three media such as newspaper/magazine, radio, and television at least once in a week (*yes, no*); maternal age in years (<20, 20-29, ≥30); gap between husband age and wife age in years (*non-positive*, 1-5, 6-10, >10); frequency of reasons for wife beating justification out of common five reasons (*not at all*, 1-2, 3-5); wealth index (*poor, middle, rich*); birth order for the index pregnancy for which ANC visits have been recorded (1, 2, 3, ≥4).

2.5.3 Results

To analyze the number of ANC visits taken by a woman during her pregnancy period, 4,427 women had been considered after adjusting for missing values. Descriptive statistics such as the mean, standard deviation, minimum and maximum for the outcome variable had been computed from the data. These statistics were found to be 2.78, 2.56, 0 and 20, respectively. The percent distribution of the number of ANC visits is shown in Figure 2.4. From the Figure 2.4, it is evident that the percentage of observed zero counts (21.28%) was much higher than the expected (6.20%) with respect to the mean of the Poisson distribution.

It was observed from the count data that the number of ANC visits in Bangladesh arises from mixture of two unobserved populations with proportions 0.52 and 0.48 in the absence of covariates. Therefore, it is reasonable to fit a Poisson-Poisson mixture model for the data set. Moreover, in order to draw inferences regarding the overall exposure effects on marginal mean, it is rational to fit MPois-Pois model, a marginally-specified mean models for mixtures of two count distributions. From Table 2.3, it is depicted that $\hat{\pi}^* = 0.610$. At first we fitted the MPois-Pois model, the negative binomial model and the Poisson model. Then, the AIC criterion had been employed to find the best fitted model. The AIC value of the fitted models for MPois-Pois, negative binomial and Poisson were found as 17922,

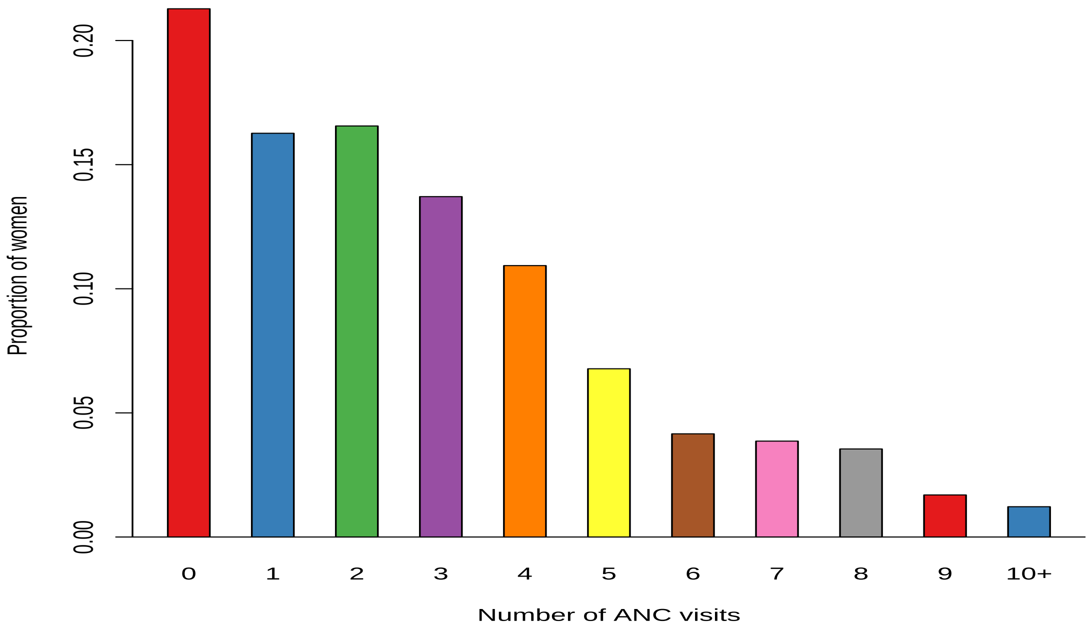


Figure 2.4: Distribution of number of ANC visits during pregnancy, BDHS 2014

18194, and 19192, respectively. Therefore, the MPois-Pois model provides the best fit to the given data set.

Table 2.3: The estimated marginal parameters ($\hat{\beta}$), standard errors (SE), and p-values under MPois-Pois mixture model on number of ANC visits, BDHS 2014

Variable name	Estimates	SE	z-value	p-value	IRR
Marginalized Poisson-Poisson model:					
Intercept	0.342	0.063	5.437	<0.001	–
Level of education					
No education (ref)					
Primary	0.279	0.057	4.899	<0.001	1.322
Secondary	0.490	0.056	8.735	<0.001	1.663
Higher	0.716	0.062	11.627	<0.001	2.047
Area of residence					
Rural (ref)					
Urban	0.160	0.026	6.148	<0.001	1.174
Media exposure					
No (ref)					
Yes	0.191	0.031	6.093	<0.001	1.210
Maternal age (years)					
<20	-0.067	0.032	-2.068	0.039	0.936
20-29 (ref)					
≥30	0.089	0.038	2.374	0.018	1.093
Gap between husband age and wife age (years)					
Non-positive	0.146	0.108	1.347	0.178	1.157
1-5 (ref)					
6-10	-0.004	0.028	-0.143	0.886	0.996
>10	-0.016	0.031	-0.521	0.603	0.984
Number of reasons wife beating justified					
Not at all (ref)					
1-2	-0.015	0.032	-0.455	0.649	0.986
3-5	-0.062	0.047	-1.305	0.192	0.940
Wealth index					
Poor (ref)					
Middle	0.121	0.038	3.168	0.002	1.128
Rich	0.328	0.043	7.712	<0.001	1.388
Birth Order					
1 (ref)					
2	-0.039	0.031	-1.254	0.210	0.961
3	-0.120	0.043	-2.782	0.005	0.887
≥4	-0.309	0.058	-5.355	<0.001	0.734
Model for Component-1:					
Intercept	-0.962	0.169	-5.701	<0.001	-
Level of education					
No education (ref)					
Primary	0.493	0.142	3.464	<0.001	-
Secondary	0.939	0.142	6.607	<0.001	-
Higher	1.266	0.153	8.301	<0.001	-
Area of residence					
Rural (ref)					
Urban	0.198	0.047	4.227	<0.001	-
Media exposure					
No (ref)					
Yes	0.327	0.057	5.722	<0.001	-
Maternal age (years)					
<20	-0.169	0.058	-2.893	0.004	-
20-29 (ref)					
≥30	0.354	0.097	3.660	<0.001	-
Gap between husband age and wife age (years)					
Non-positive	-0.076	0.202	-0.376	0.707	-
1-5 (ref)					
6-10	0.063	0.052	1.229	0.219	-
>10	0.136	0.058	2.324	0.020	-

Continued...Table 2.3

Variable name	Estimates	SE	z-value	p-value	IRR
Number of reasons wife beating justified					
Not at all (ref)					
1-2	-0.048	0.058	-0.829	0.407	-
3-5	-0.153	0.089	-1.722	0.085	-
Wealth index					
Poor (ref)					
Middle	0.421	0.074	5.667	<0.001	-
Rich	0.698	0.080	8.673	<0.001	-
Birth Order					
1 (ref)					
2	-0.170	0.060	-2.846	0.004	-
3	-0.268	0.078	-3.453	0.001	-
$\geq$4	-0.717	0.122	-5.871	<0.001	-
Mixing Proportion					
$\hat{\pi}$	0.610	0.043	14.339	<0.001	-
AIC	17922				

2.6 Conclusion

When analyzing count data, it is necessary to determine if the data originate from a population or from a mixture of populations in order to draw valid inferences. If the target population consists of a mixture of populations and if the augmentation of zeros arises for this reason, the latent class parameters for the regression model can be estimated by fitting a mixture model. In case of mixture model setup, however, it is not possible to estimate the regression parameters for modeling marginalized means. Hence, inference regarding the exposure effects in terms of IRR for the population-wide parameters cannot be obtained from such a model. This problem can be solved by developing a marginalized (marginalization over the subpopulations) mixture model for drawing valid inference.

Extensive simulation studies were carried out to investigate the MPois-Pois model under cross-sectional setup. According to the simulation studies, the MPois-Pois mixture model offers minimal MSE and confidence interval coverages near to the nominal levels when the true model is specified; and misspecification of the model increases MSE and under-estimates the coverage probability.

Real data set extracted from the Bangladesh Demographic and Health Survey (BDHS), 2014 has been utilized for the application of the MPois-Pois model under a cross-sectional setup. It is observed that the MPois-Pois model is the best fitted model than the other count models.

Chapter 3

Marginalized Mixture Models: Zero-Inflated Clustered Count Data

It may happen in practice that zero-inflated clustered count data or mixture of two-component clustered count data may arise in a variety of contexts (Long et al., 2015; Wang et al., 2007; Min and Agresti, 2005; Yau et al., 2003; Wang et al., 2002; Hall, 2000). There are several modeling approach in the existing literature for analyzing zero-inflated clustered count data. Under existing model of zero-inflated clustered count data, Hall (2000) suggested a ZIP model modification that includes random effects in the Poisson process to take within-cluster correlation into consideration; Yau et al. (2003) proposed a zero-inflated negative binomial (ZINB) regression model in order to account for excess zeros as well as over-dispersion with independent random effects in each process; and Min and Agresti (2005) proposed hurdle model as a two-part model with independent random effects in each part.

Numerous studies using the existing mixture modeling approach (mixture of 'not-at-risk' and 'at-risk' classes) for analyzing zero-inflated clustered count data have clearly shown the importance of inferences regarding the marginal mean, and the researchers have developed models in order to do so (Tabb et al., 2016; Long et al., 2015; Kassahun et al., 2014; Lee et al., 2011; Hall and Zhang, 2004). Making a covariate-adjusted inference for the marginal mean (marginalization over the subpopulations) of an exposure effect was the researchers' aim in such cases.

In the previous chapters, we have already provided the reasons of considering Poisson-Poisson mixture instead of existing models for analyzing zero-inflated count data. In analyzing zero-inflated clustered data, in this thesis, it is assumed that an observation unit in a cluster belongs to one of two latent populations each with Poisson distribution. In such data, unobserved heterogeneity arises due to the presence of mixture in the dgp (Wang et al., 2007) and existence of correlation among the observations (Ridout et al., 1999) in a cluster. Although unobserved heterogeneity can be controlled by utilizing a negative binomial model or employing a mixed modeling approach in the standard count models, it is not possible to remove all the heterogeneity that have been considered in analyzing zero-inflated clustered data by fitting such models.

Wang et al. (2007, 2002) have proposed a Poisson-Poisson mixture model for analyzing clustered count data, where counts arise from two latent classes. Like cross-section setup, the main limitation of such model is that the interpretation of the estimates is based on the latent classes, and hence the inference of the parameters regarding marginal mean cannot be made from the mixture model framework of clustered data.

To the best of our knowledge, the marginalized Poisson-Poisson model for analyzing clustered data has yet not been studied by any researcher. This motivates to develop a marginalized mean model that would be useful for analyzing clustered count data arising from a mixture of two latent subpopulations. It can be accomplished by extending cross-sectional MPois-Pois (Benecha et al., 2017) model in clustered data setup. To model the marginal mean under standard mixture model setup for analyzing zero-inflated clustered data, following Long et al. (2015) we proposed an extension of MPois-Pois model (Benecha et al., 2017) under the framework of ML estimation. The proposed model is refer to as

Random Effects Marginalized Poisson-Poisson (REMPois-Pois) mixture model.

To develop REMPois-Pois model, we assume that (i) data are collected from a population is divided into a number of clusters; (ii) population observations in each cluster arises from one of the two latent subpopulations with remarkable differences in means; (iii) the random effects are incorporated into the linear predictors of the model for marginal mean and component-1 mean as random intercepts to control the unobserved heterogeneity in the data.

3.1 Random Effects Marginalized Poisson-Poisson Mixture Model

Suppose that K clusters are selected randomly from a population and the n_i observations are selected from the ith cluster. The sample size is then expressed as $n = \sum_{i=1}^{K} n_i$. Let $Y_{ij}(i = 1, \cdots, K; j = 1, \cdots, n_i)$ be the count variable of interest for the jth individual in the ith cluster. The Poisson-Poisson mixture distribution and its marginalized model for cross-sectional setup were discussed in chapter 2. In the FMM setup for clustered data, the probability distribution is known as the Poisson-Poisson mixture distribution with bivariate random effects (Wang et al., 2007, 2002). It is assumed that the ith cluster-population is divided into two latent subpopulations (component-1 and component-2) and for a given random effect, the observations of component-1 follow independent Poisson distributions with low means and that of component-2 follow independent Poisson distributions with larger means. Let $\mu^*_{1,ij} = E\big[Y_{ij}|c_i\big]$ and $\mu^*_{2,ij} = E\big[Y_{ij}|d_i\big]$, where c_i and d_i are the random effects term corresponding to the component-1 mean ($\mu^*_{1,ij}$) and the component-2 mean ($\mu^*_{2,ij}$), respectively. Let $b_i^T = (c_i, d_i) \sim N_2(\mathbf{0}, \Sigma^*)$, where Σ^* is a 2×2 covariance matrix with diagonal elements σ_c^2, σ_d^2 and off-diagonal element $\rho^{**}\sigma_c\sigma_d$. Following Eq.(2.17), the regression models for Poisson-Poisson mixture random variables for analyzing clustered data can be written as

$$\log(\mu_{1,ij}^*) = \boldsymbol{x}_{ij}'\boldsymbol{\alpha} + c_i; \quad \log(\mu_{2,ij}^*) = \boldsymbol{x}_{ij}'\boldsymbol{\gamma} + d_i; \quad \text{logit}(\pi^*) = \tau; \tag{3.1}$$

where $\boldsymbol{x}_{ij}$ is the set of covariates for the jth individual in the ith cluster used in both the models for component-1 and component-2; $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are $p \times 1$ vector of and parameters corresponding to component-1 and component-2, respectively, and $\infty < \tau < \infty$. Instead of modeling the latent class Poisson means $\mu_{1,ij}^*$ and $\mu_{2,ij}^*$, following Long et al. (2015), an extension of MPois-Pois model (Benecha et al., 2017) called REMPois-Pois model for analyzing clustered data is proposed in this study. The proposed model can directly models the cluster-specific overall mean $E\big[Y_{ij}|u_i,\big] = \mu_{ij}^*$ through the following link functions

$$\log(\mu_{ij}^*) = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + u_i; \quad \log(\mu_{1,ij}^*) = \boldsymbol{x}_{ij}'\boldsymbol{\alpha} + c_i; \quad \text{logit}(\pi^*) = \tau, \tag{3.2}$$

where u_i and c_i are the random effects term corresponding to the cluster-specific overall mean and the conditional mean of component-1, respectively. The random effects $v_i^T = (u_i, c_i)$ and are assumed to be distributed as $N_2(\mathbf{0}, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_u^2 & \rho^*\sigma_u\sigma_c \\ \rho^*\sigma_u\sigma_c & \sigma_c^2 \end{pmatrix}$$

Under the marginalized model as in Eq.(3.2), the cluster-specific overall mean and variance of Y_{ij} can be written as

$$\mu_{ij}^* = \pi^*\mu_{1,ij}^* + (1-\pi^*)\mu_{2,ij}^*, \tag{3.3}$$

and

$$\text{Var}\big[Y_{ij}|u_i, c_i\big] = \mu_{ij}^* + \left[\frac{\pi^*}{1-\pi^*}\right]\big(\mu_{ij}^* - \mu_{1,ij}^*\big)^2, \tag{3.4}$$

respectively.

3.1.1 Likelihood Function

Let $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{in_i})'$ be the $n_i \times 1$ response vector for ith cluster $(i = 1, \cdots, K)$. The observations in the ith cluster are conditionally independent. Therefore, the conditional distribution of $\boldsymbol{y}_i$ given the random effects u_i, c_i is given as follows

$$f(\boldsymbol{y}_i|u_i, c_i; \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = \prod_{j=1}^{n_i} f(y_{ij}|u_i, c_i; \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau),$$

where following Benecha et al. (2017), the probability distribution function, $f(y_{ij})$ can be written as

$$f(y_{ij}|u_i, c_i; \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = \pi^* \frac{e^{-\mu^*_{1,ij}} {\mu^*_{1,ij}}^{y_{ij}}}{y_{ij}!} + (1-\pi^*) \frac{e^{-\dfrac{\mu^*_{ij} - \pi^*\mu_{1,ij}}{1-\pi^*}} \left(\dfrac{\mu^*_{ij} - \pi^*\mu_{1,ij}}{1-\pi^*}\right)^{y_{ij}}}{y_{ij}!}. \quad (3.5)$$

The probability distribution given in Eq.(3.5) is named as REMPois-Pois distribution. The contribution of ith cluster in the (unconditional) likelihood function is then

$$\begin{aligned}
\mathrm{L}_i(\boldsymbol{\beta}, \sigma_u, \boldsymbol{\alpha}, \sigma_c, \tau, \rho^*|\boldsymbol{y}_i) =& f(\boldsymbol{y}_i|\boldsymbol{\beta}, \sigma_u, \boldsymbol{\alpha}, \sigma_c, \tau, \rho^*) \\
=& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\boldsymbol{y}_i; u_i, c_i|\boldsymbol{\beta}, \sigma_u, \boldsymbol{\alpha}, \sigma_c, \tau, \rho^*) du_i dc_i \\
=& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\boldsymbol{y}_i|u_i, c_i; \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) f(u_i, c_i|\sigma_u, \sigma_c, \rho^*) du_i dc_i \\
=& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f(y_{ij}|u_i, c_i; \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) f(u_i, c_i|\sigma_u, \sigma_c, \rho^*) du_i dc_i.
\end{aligned}$$

For simplicity, let us consider $\boldsymbol{\delta} = (\boldsymbol{\beta}', \sigma_u, \boldsymbol{\alpha}', \sigma_c, \tau, \rho^*)'$. Since the clusters are independent, the overall likelihood function can be computed as

$$\begin{aligned}
\mathrm{L}(\boldsymbol{\delta}|\boldsymbol{y}_1, \ldots \boldsymbol{y}_K) =& \prod_{i=1}^{K} \mathrm{L}_i(\boldsymbol{\beta}, \sigma_u, \boldsymbol{\alpha}, \sigma_c, \tau, \rho^*|\boldsymbol{y}_i) \\
=& \prod_{i=1}^{K} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f(y_{ij}|u_i, c_i; \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) f(u_i, c_i|\sigma_u, \sigma_c, \rho^*) du_i dc_i. \quad (3.6)
\end{aligned}$$

The log-likelihood function is then written as follows

$$\ell(\boldsymbol{\delta}|\boldsymbol{y}_1, \ldots \boldsymbol{y}_K) = \sum_{i=1}^{K} \log\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f(y_{ij}|u_i, c_i; \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) \phi(u_i, c_i|\sigma_u, \sigma_c, \rho^*) du_i dc_i\right]. \quad (3.7)$$

Maximization of the log-likelihood function (3.7) is naturally complicated because the integrals inside the function cannot be expressed in closed form. As the joint probability distribution of the random effects for the overall-means and the component-1 means is bivariate normal, i.e., $v_i^T = (u_i, c_i) \sim N_2(\mathbf{0}, \Sigma)$, this complication could be overcome by employing Gauss–Hermite quadrature (GHQ) to approximate the integrals inside the logarithm of Eq.(3.7) over the random effects (Hall, 2000). After then a reasonable initial guess of the parameters were considered for getting a converged sequence of estimates. For this purpose, following Benecha et al. (2017), we have utilized estimates from random effects Poisson model for the initial values of $\boldsymbol{\beta}$ and σ_u, and estimates of parameters for model of component-1 from random effects two-component Poisson mixture model for the initial values of $\pi^*, \boldsymbol{\alpha}$ and σ_c.

The expression of the approximate log-likelihood function of REMPois-Pois model can be obtained by using the GHQ technique (Hall, 2000; Pinheiro and Bates, 1995), which is provided in Appendix B.1. Using the link functions of Eq.(3.2), the log-likelihood function of Eq.(3.7) can be then approximated by

$$\begin{aligned}
&\sum_{i=1}^{K} \log \left[\sum_{l_1=1}^{m} \sum_{l_2=1}^{m} \frac{w_{l_1}^u w_{l_2}^c \sqrt{1-\rho^{*2}} e^{2\rho^* q_{l_1}^u q_{l_2}^c}}{\pi} \prod_{j=1}^{n_i} \left\{ \frac{1}{y_{ij}! \left(1+e^{\tau}\right)} \right. \right. \\
&\left(e^{\left[\tau + y_{ij}(\boldsymbol{x}_{ij}'\boldsymbol{\alpha} + \sqrt{2(1-\rho^{*2})}\sigma_c q_{l_2}^c) - e^{\boldsymbol{x}_{ij}'\boldsymbol{\alpha} + \sqrt{2(1-\rho^{*2})}\sigma_c q_{l_2}^c}\right]} \right. \\
&+ e^{-\left[\left(1+e^{\tau}\right) e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \sqrt{2(1-\rho^{*2})}\sigma_u q_{l_1}^u} - e^{\tau + \boldsymbol{x}_{ij}'\boldsymbol{\alpha} + \sqrt{2(1-\rho^{*2})}\sigma_c q_{l_2}^c}\right]} \\
&\left. \left. \left. \left[\left(1+e^{\tau}\right) e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \sqrt{2(1-\rho^{*2})}\sigma_u q_{l_1}^u} - e^{\tau + \boldsymbol{x}_{ij}'\boldsymbol{\alpha} + \sqrt{2(1-\rho^{*2})}\sigma_c q_{l_2}^c}\right]^{y_{ij}} \right) \right\} \right].
\end{aligned} \tag{3.8}$$

If the random effects for the overall-means and the component-1 means are assumed independent ($\rho^* = 0$), the log-likelihood function can be expressed by the following approximation

$$\sum_{i=1}^{K} \log \Bigg[\sum_{l_1=1}^{m} \sum_{l_2=1}^{m} \frac{w_{l_1}^u w_{l_2}^c}{\pi} \prod_{j=1}^{n_i} \Bigg\{ \frac{1}{y_{ij}!\left(1+e^{\tau}\right)} \Bigg(e^{\left[\tau + y_{ij}(\boldsymbol{x}_{ij}'\boldsymbol{\alpha} + \sqrt{2}\sigma_c q_{l_2}^c) - e^{\boldsymbol{x}_{ij}'\boldsymbol{\alpha} + \sqrt{2}\sigma_c q_{l_2}^c}\right]} + e^{-\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \sqrt{2}\sigma_u q_{l_1}^u} - e^{\tau + \boldsymbol{x}_{ij}'\boldsymbol{\alpha} + \sqrt{2}\sigma_c q_{l_2}^c}\right]} \left[\left(1+e^{\tau}\right) e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \sqrt{2}\sigma_u q_{l_1}^u} - e^{\tau + \boldsymbol{x}_{ij}'\boldsymbol{\alpha} + \sqrt{2}\sigma_c q_{l_2}^c}\right]^{y_{ij}} \Bigg) \Bigg\} \Bigg]. \tag{3.9}$$

As $-1 \leq \rho^* \leq 1$ is a restricted range parameter, we have encountered a problem in directly estimating ρ^* by quasi-Newton optimization. This problem can be overcome by using the cumulative distribution function (cdf) of exponential distribution. Let us consider that the parameter $-\infty < \rho^*_{unres} < \infty$ will be plugged-in in the optimization method. Then it can be shown the quantity, $\text{sign}(\rho^*_{unres}) F_S(\mid \rho^*_{unres} \mid)$ will produce value between -1 and +1, where S is a exponential random variable with mean 1 with cdf $F(\cdot)$. Therefore, the problem has been addressed in this study by considering the transformation as follows

$$\rho^* = \text{sign}(\rho^*_{unres})\left(1 - e^{-|\rho^*_{unres}|}\right).$$

To obtain the MLE of ρ^*, first we have estimated the unrestricted range parameter $-\infty < \rho^*_{unres} < \infty$ by quasi-Newton method and then the restricted range parameter had been estimated as

$$\hat{\rho}^* = \text{sign}(\hat{\rho}^*_{unres})\left(1 - e^{-|\hat{\rho}^*_{unres}|}\right).$$

3.1.2 Score Function

The score function can be defined from the approximated log-likelihood function of REMPois-Pois model given in Eq.(3.9) as

$$U(\boldsymbol{\delta}) = \left[\frac{\partial l(\boldsymbol{\delta})}{\partial \tau}, \frac{\partial l(\boldsymbol{\delta})}{\partial \boldsymbol{\alpha}'}, \frac{\partial l(\boldsymbol{\delta})}{\partial \sigma_c}, \frac{\partial l(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}'}, \frac{\partial l(\boldsymbol{\delta})}{\partial \sigma_u}\right]' = \left[U_1, U_2', U_3, U_4', U_5\right]'$$

The expression for the elements of score function was given in the Appendix B.2 . Using these elements, the maximum likelihood estimating equations for $\boldsymbol{\delta}$ of the REMPois-Pois model can be formed as

$$U(\boldsymbol{\delta}) = \mathbf{0}_{(2p+3)\times 1}. \tag{3.10}$$

The solution of Eq.(3.10) can be obtained by Newton's method. At the rth step, Newton's method updates the values of the parameters as

$$\boldsymbol{\delta}^{(r)} = \boldsymbol{\delta}^{(r-1)} + \left(I(\boldsymbol{\delta}^{(r-1)})\right)^{-1} U(\boldsymbol{\delta}^{(r-1)}), r = 1, 2, \cdots, \tag{3.11}$$

for $r = 1, 2, \cdots$ until convergence. In Eq.(3.11), $I(\cdot)$ is a $(2p+3)\times(2p+3)$ matrix of observed information obtained from negative of Hessian. Although we can derive the expression of Hessian for the REMPois-Pois model, we have refrained providing the complex expressions of Hessian.

With carefully chosen starting values, the MLE of the parameters $(\hat{\boldsymbol{\delta}})$ are the solution of Eq.(3.10) and can be obtained by the use of quasi-Newton optimization method or the Newton's method. For the proposed REMPois-Pois model, the quasi-Newton optimization technique was implemented by R 'optim' function. Starting values for marginal parameters $(\boldsymbol{\beta}, \sigma_u)$ are the fitted values of random effects Poisson model. Also, the initial values of $(\tau, \boldsymbol{\alpha}', \sigma_c)$ can be obtained by fitting the random effects Poisson-Poisson mixture model(Wang et al., 2007) by applying the EM algorithm with GHQ (Hall, 2000). The EM algorithm for estimating parameters of Poisson-Poisson mixture regression model with random effects is given in the Appendix B.3. This method of initialization was recommended by Benecha et al. (2017) for obtaining the MLE of parameters of marginalized Poisson-Poisson mixture model in cross-sectional setup.

Let $I^{ll}(\boldsymbol{\delta}), l = 1, 2, \ldots, (2p + 3)$ be the (l, l)th element of the inverse of information matrix. Then the standard error of the lth component of the estimator of the parameter vector $\boldsymbol{\delta}$ is given by

$$\mathrm{se}(\hat{\boldsymbol{\delta}}_l) = \sqrt{I^{ll}(\boldsymbol{\delta})}. \tag{3.12}$$

For lth parameter, $l = 1, 2, \cdots, (2p+3)$, the asymptotic behavior of the estimator can be expressed as

$$\sqrt{n}\big(\hat{\boldsymbol{\delta}}_l - \boldsymbol{\delta}_l\big) \sim N\Big(0, \mathrm{se}(\hat{\boldsymbol{\delta}}_l)^2\Big).$$

3.1.3 Intraclass Correlations

The term intraclass correlation coefficient (ICC) is commonly used in measuring the degree of resemblance of units in the same cluster (Oliveira et al., 2016). The ICC is computed by marginalization over the clusters and thus it is regarded as a population-averaged (PA) characteristics. In analyzing clustered data, cluster-specific characteristics are usually introduced by incorporating the random effects into the model for taking into account the ICC. Although it is of interest to estimate the fixed effects parameters along with the random effects parameters, a measure of ICC is also important in such instance. Under regression setup, the population averaged (PA) characteristics of Y_{ij} for the proposed REMPois-Pois model can be determined by computing the unconditional mean, variance, covariance and intraclass correlation. To compute the PA characteristics, let us define two standard normal random variables as $U_i^* = \sigma_u^{-1} U_i$ and $C_i^* = \sigma_c^{-1} C_i$. Therefore, the moment generating function of U_i^* and C_i^*can be defined as $M_{U_i^*}(t) = e^{\frac{1}{2}t^2} = M_{C_i^*}(t)$.

Unconditional Mean and Variance

The expression for unconditional mean and variance of Y_{ij} can be derived as

$$\mu_{ij} = E(Y_{ij}) = E\Big[E\big(Y_{ij}|u_i\big)\Big] = E\Big[e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sigma_u U_i^*}\Big] = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}} E\Big[e^{\sigma_u U_i^*}\Big] = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}} M_{U_i^*}(\sigma_u) = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\frac{1}{2}\sigma_u^2}, \tag{3.13}$$

and

$$\begin{aligned}
\text{Var}(Y_{ij}) =& \text{Var}\Big[E\big(Y_{ij}|u_i\big)\Big] + E\Big[\text{Var}\big(Y_{ij}|u_i, c_i\big)\Big] \\
=& \text{Var}\Big[e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sigma_u U_i^*}\Big] + E\Big[e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sigma_u U_i^*} + e^{\tau}\big(e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sigma_u U_i^*} - e^{\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sigma_c C_i^*}\big)^2\Big] \\
=& \Big\{E\Big[e^{2\big(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sigma_u U_i^*\big)}\Big] - \mu_{ij}^2\Big\} + \Big\{\mu_{ij} + e^{\tau}\Big(E\Big[e^{2\big(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sigma_u U_i^*\big)}\Big] + E\Big[e^{2\big(\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sigma_c C_i^*\big)}\Big] \\
& - 2E\Big[e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sigma_u U_i^*+\sigma_c C_i^*}\Big]\Big)\Big\} \\
& = (1+e^{\tau})e^{2\big(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sigma_u^2\big)} + \mu_{ij}(1-\mu_{ij}) + e^{\tau}\Big\{e^{2\big(\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sigma_c^2\big)} - 2\mu_{ij}\mu_{1,ij}e^{\rho^*\sigma_u\sigma_c}\Big\}.
\end{aligned} \tag{3.14}$$

Unconditional Covariances

The unconditional covariances for the jth and lth individuals in the ith cluster $i = 1, \cdots, K, j = 1, \cdots, n_i, l = 1, \cdots, n_i, j \neq l$ can be obtained as

$$\begin{aligned}
\text{Cov}\Big(Y_{ij}, Y_{il}\Big) =& \text{Cov}\Big[E\big(Y_{ij}|u_i\big), E\big(Y_{il}|u_i\big)\Big] + E\Big[\text{Cov}\big(Y_{ij}, Y_{il}|u_i\big)\Big] \\
=& \text{Cov}\Big[e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sigma_u U_i^*}, e^{\boldsymbol{x}'_{il}\boldsymbol{\beta}+\sigma_u U_i^*}\Big] \\
=& E\Big[e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sigma_u U_i^*} e^{\boldsymbol{x}'_{il}\boldsymbol{\beta}+\sigma_u U_i^*}\Big] - \mu_{ij}\mu_{il} \\
=& e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{x}'_{il}\boldsymbol{\beta}} E\Big[e^{2\sigma_u U_i^*}\Big] - \mu_{ij}\mu_{il} \\
=& e^{\big(\boldsymbol{x}'_{ij}+\boldsymbol{x}'_{il}\big)\boldsymbol{\beta}}\Big(e^{2\sigma_u^2} - e^{\sigma_u^2}\Big).
\end{aligned} \tag{3.15}$$

The correlation coefficient of jth and lth individuals in the ith cluster under the proposed REMPois-Pois model can be computed from the expression of unconditional variance and covariance as follows

$$\rho_{ijl} = \frac{\text{Cov}\Big(Y_{ij}, Y_{il}\Big)}{\sqrt{\text{Var}(Y_{ij})\text{Var}(Y_{il})}}, i = 1, \cdots, K, j = 1, \cdots, n_i - 1, l = j+1, \cdots, n_i,$$

where $\text{Cov}\Big(Y_{ij}, Y_{il}\Big)$ is the unconditional covariance as given in Eq.(3.15) and $\text{Var}(Y_{ij})$ is the unconditional variance as given in Eq.(3.14). The intra-cluster correlation is then computed as

$$\rho = \sum_{i=1}^{K} \sum_{j=1}^{n_i-1} \sum_{l=j+1}^{n_i} \rho_{ijl} \Big/ \sum_{i=1}^{K} \binom{n_i}{2}. \tag{3.16}$$

The intra-cluster correlation (ρ) as shown in Eq.(3.16) can be estimated by using the estimates of unconditional covariance ($\hat{\text{Cov}}(Y_{ij}, Y_{il})$) and the estimates of unconditional variance ($\hat{\text{Var}}(Y_{ij})$). From Eq.(3.14) and Eq.(3.15), one can easily estimate the unconditional variances and unconditional covariances by using the estimates of fixed effects regression parameters ($\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$), the estimates of random effects parameters ($\hat{\sigma}_u$ and $\hat{\sigma}_c$), the estimates of parameters for mixing proportion ($\hat{\tau}$), and the estimates of correlation coefficient between the random effects ($\hat{\rho}^*$).

3.2 Simulation Study

In the simulation studies, it was assumed that both model for conditional marginalized mean (marginalized over the subpopulations) and model for conditional component-1 mean given in Eq.(3.2) were influenced by the same set of known covariates $\boldsymbol{x}_{ij} = (x_{0ij}, x_{1ij}, x_{2ij}, \cdots, x_{(p-1)ij})'$ with $x_{0ij} = 1$ for the ith response. In order to generate zero-inflated count in clustered data setup, the parameters were chosen in such a way that $0 \leq \mu_{1,ij} << \mu_{2,ij} < \infty$ would be observed. The number of cluster is denoted by K and for the ith cluster ($i = 1, \cdots, K$), the cluster size is denoted by n_i. The sample size, n can be expressed as, $n = \sum_{i=1}^{K} n_i$. Note that the data were generated using a two components random effects Poisson mixture model using probability π^* and we were interested in the inference regarding the marginal (marginalization over the subpopulations) means.

3.2.1 The situation when $\rho^* = 0$

In this situation, the random effects of marginal mean, U and the random effects of component-1 mean, C were considered independent for simplicity. Extensive simulation studies were conducted to investigate the performance of the proposed REMPois-Pois model for different number of clusters and varying cluster sizes. For instance, we had considered $K = 50$, 100, 200 and $n_i = 5$, 15, 30 while keeping same size for all the clusters for a particular value of K i.e., $n_1 = \cdots = n_K$. To obtain marginal inference from random effects Poisson-Poisson mixture distribution, the zero-inflated data for a sample of size n had been generated by the following steps where different mixing probabilities such as $\pi^* = 0.50, 0.70, 0.90$ had been used for varying the proportion of zeros in the data.

1. K random variables, $U_1, \cdots, U_K$ from $N(0, \sigma_u^2)$ and K random variables, $C_1, \cdots, C_K$ from $N(0, \sigma_c^2)$ had been generated and each value of both variables was then repeated n_i times $(i = 1, \cdots, K)$ to make up the n values of U and C.

2. The covariate x_{1ij} were generated from unif(0,1), and the covariate x_{2ij} were generated from Bernoulli(0.40), $i = 1, \cdots, K, j = 1, \cdots, n_i$.

3. In Eq.(3.2), suitable values of the fixed effect regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)' = (0.20, 0.40, 0.30)'$ and random effects parameter $\sigma_u = 0.40$ were used to compute the marginal means $\mu^*_{ij}, i = 1, \cdots, K, j = 1, \cdots, n_i$ and suitable values of the fixed effect regression parameters $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)' = (-1.50, 0.30, 0.20)'$ and random ef-

fects parameter $\sigma_c = 0.30$ were used to compute the component-1 means $\mu^*_{1,ij}, i = 1, \cdots, K, j = 1, \cdots, n_i$.

4. The component-2 means ($\mu^*_{2,ij}$) were then computed using Eq.(3.3). If $\mu^*_{2,ij} < 0$ was observed in this process, the set of values $\boldsymbol{x}_{ij} = (x_{ij0}, x_{ij1}, x_{ij2})'$ was replaced by randomly selected set of values $\boldsymbol{x}_{il} = (x_{il0}, x_{il1}, x_{il2})'$ for which $\mu^*_{2,il} \geq 0$ had been observed, $j \neq l$.

5. Binary observations $d_{ij}, i = 1, \cdots, K, j = 1, \cdots, n_i$ were generated using $D \sim$Bernoulli(π^*).

6. Under REMPois-Pois setup, the n zero-inflated clustered count observations were then obtained by

$$y_{ij} = \begin{cases} y_{1,ij} \sim \text{Pois}(\mu^*_{1,ij}), & \text{if} \quad d_{ij} = 1 \\ y_{2,ij} \sim \text{Pois}(\mu^*_{2,ij}), & \text{if} \quad d_{ij} = 0. \end{cases}$$

for $i = 1, \cdots, K, j = 1, \cdots, n_i$.

7. For different values of K, n_i and for given values of $\boldsymbol{\beta}, \sigma_u, \boldsymbol{\alpha}, \sigma_c$, various proportion of zeros have been generated. For instances, data contain approximately 40% zeros for $\pi^* = 0.50$ ($\tau = 0.000$), approximately 50% zeros for $\pi^* = 0.70$ ($\tau = 0.8473$), and approximately 65% zeros for $\pi^* = 0.90$ ($\tau = 2.1972$).

The simulation was repeated 1000 times for each setup. We then estimated the regression parameters, the random effects parameters and all other nuisance parameters by using maximum likelihood (ML) approach. In order to investigate the performance of ML estimates, we have computed the bias, standard error and the coverage probability (Cov.Pr.). The biases were computed from the differences between simulated means (SM) and the true values for each of the parameters. Two types of standard errors such as estimated standard errors (ESE) and simulated standard errors (SSE) were also computed to investigate the properties of the estimators. The SM, ESE and SSE for a estimator $\hat{\delta}$ were calculated similarly as in Eq.(2.29). The proportion of convergences (Conv.Prop.) in fitting the REMPois-Pois model were computed for all the setups.

The SM, simulated mean of mixing proportion (SMP), Bias, ESE, SSE, Cov.Pr. and Conv.Prop. were given in Table 3.1 for mixing proportion $\pi^* = 0.50$. From Table 3.1, it is clear that the estimates of marginal parameters for fixed effects had minimal amount of biases for all the settings. However, random effect parameter had highest bias for $K = 50$ and $n_i = 5$, the bias decreases with increasing the cluster size and/or with increasing the number of clusters. For example, the amount of bias for $(\beta_0, \beta_1, \beta_2)$ was $(-0.005, 0.004, -0.001)$ when $n_i = 5$, it was $(0.001, -0.004, 0.001)$ if $n_i = 15$ and $(0.006, -0.002, 0.000)$ when $n_i = 30$ for $K = 50$. These figures were $(-0.003, 0.007, -0.010)$, $(0.004, -0.003, -0.002)$ and $(0.007, 0.001, -0.003)$ when $K = 100$; $(0.002, 0.001, -0.003)$, $(-0.003, 0.000, 0.001)$and $(0.006, -0.002, 0.000)$ if $K = 200$ for $n_i = 5, 15$ and $n_i = 30$, respectively. The amount of bias of σ_u were -0.034, -0.008 and -0.007 when $n_i = 5$, 15 and 30, respectively for $K = 50$. These amounts were -0.026, -0.001 and -0.008 for $n_i = 5$, 15 and 30, respectively if $K = 100$; and -0.010, -0.006 and -0.013 for $n_i = 5$, 15 and 30, respectively when $K = 200$.

The amount of bias also decreases with increasing the cluster size and/or with increasing the number of clusters for most of the nuisance parameters. For example, the amount of bias of $(\alpha_0, \alpha_1, \alpha_2)$ was $(0.024, -0.055, -0.014)$ when $n_i = 5$, $(0.002, -0.026, -0.006)$ if $n_i = 15$ and it was $(-0.007, -0.012, 0.003)$ when $n_i = 30$ for $K = 50$. These figures were $(-0.009, 0.046, -0.021)$, $(-0.014, 0.002, -0.010)$ and $(0.000, 0.003, -0.010)$ for $n_i = 5$, 15 and 30, respectively when $K = 100$; $(-0.006, 0.002, -0.015)$, $(-0.005, -0.002, -0.003)$ and $(0.004, 0.001, 0.002)$ for $n_i = 5$, 15 and 30, respectively if $K = 200$.

The amount of bias of σ_c was 0.073, -0.049 and -0.033 when $n_i = 5$, 15 and 30, respectively for $K = 50$. These amounts were 0.003, -0.05 and -0.008 for $n_i = 5$, 15 and 30, respectively if $K = 100$; and -0.008, -0.024 and 0.013 for $n_i = 5$, 15 and 30, respectively when $K = 200$.

Table 3.1: Simulated mean (SM), simulated mean of mixing proportion (SMP), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) for estimating Marginal Parameters: $(\boldsymbol{\beta}, \sigma_u)$; Component-1 Parameters $(\boldsymbol{\alpha}, \sigma_c)$ and τ for mixing proportion $\pi^* = 0.50$ from REMPois-Pois model for different number of clusters and various cluster sizes

(K, n_i)	Conv.Prop.	Params	SM	SMP	Bias	ESE	SSE	Cov.Pr.
(50,5)		$\beta_0 = 0.20$	0.195		-0.005	0.164	0.175	94.3
		$\beta_1 = 0.40$	0.404		0.004	0.244	0.262	94.4
		$\beta_2 = 0.30$	0.299		-0.001	0.142	0.147	95.6
		$\sigma_u = 0.40$	0.366		-0.034	0.100	0.125	94.6
	0.789	$\alpha_0 = -1.50$	-1.476	0.521	0.024	0.623	0.863	95.2
		$\alpha_1 = 0.30$	0.245		-0.055	0.857	1.223	94.6
		$\alpha_2 = 0.20$	0.186		-0.014	0.518	0.623	94.4
		$\sigma_c = 0.30$	0.373		0.073	0.508	0.430	94.8
		$\tau = 0.0000$	0.082		0.082	0.232	0.254	94.2
(50,15)		$\beta_0 = 0.20$	0.201		0.001	0.095	0.105	95.4
		$\beta_1 = 0.40$	0.396		-0.004	0.123	0.124	94.2
		$\beta_2 = 0.30$	0.301		0.001	0.072	0.073	94.9
		$\sigma_u = 0.40$	0.392		-0.008	0.051	0.067	95.4
	0.846	$\alpha_0 = -1.50$	-1.498	0.510	0.002	0.356	0.363	95.5
		$\alpha_1 = 0.30$	0.274		-0.026	0.499	0.520	94.7
		$\alpha_2 = 0.20$	0.194		-0.006	0.291	0.307	94.6
		$\sigma_c = 0.30$	0.251		-0.049	0.355	0.249	97.8
		$\tau = 0.0000$	0.039		0.039	0.123	0.189	96.9
(50,30)		$\beta_0 = 0.20$	0.206		0.006	0.072	0.093	94.8
		$\beta_1 = 0.40$	0.398		-0.002	0.084	0.086	94.6
		$\beta_2 = 0.30$	0.300		0.000	0.049	0.049	95.0
		$\sigma_u = 0.40$	0.393		-0.007	0.035	0.057	95.3
	0.920	$\alpha_0 = -1.50$	-1.507	0.501	-0.007	0.261	0.274	94.9
		$\alpha_1 = 0.30$	0.288		-0.012	0.351	0.360	94.8
		$\alpha_2 = 0.20$	0.203		0.003	0.203	0.206	94.2
		$\sigma_c = 0.30$	0.267		-0.033	0.220	0.186	99.2
		$\tau = 0.0000$	0.005		0.005	0.085	0.084	95.5
(100, 5)		$\beta_0 = 0.20$	0.197		-0.003	0.114	0.121	95.3
		$\beta_1 = 0.40$	0.407		0.007	0.167	0.174	94.3
		$\beta_2 = 0.30$	0.290		-0.010	0.097	0.100	94.8
		$\sigma_u = 0.40$	0.374		-0.026	0.071	0.080	93.6
	0.859	$\alpha_0 = -1.50$	-1.509	0.513	-0.009	0.444	0.495	94.1
		$\alpha_1 = 0.30$	0.346		0.046	0.604	0.726	94.3
		$\alpha_2 = 0.20$	0.179		-0.021	0.361	0.409	94.8
		$\sigma_v = 0.30$	0.303		0.003	0.502	0.372	94.6
		$\tau = 0.0000$	0.052		0.052	0.172	0.182	94.3
(100, 15)		$\beta_0 = 0.20$	0.204		0.004	0.070	0.075	95.1
		$\beta_1 = 0.40$	0.397		-0.003	0.086	0.090	94.8
		$\beta_2 = 0.30$	0.298		-0.002	0.050	0.052	95.4
		$\sigma_u = 0.40$	0.399		-0.001	0.035	0.046	94.7
	0.938	$\alpha_0 = -1.50$	-1.514	0.501	-0.014	0.263	0.271	94.6
		$\alpha_1 = 0.30$	0.302		0.002	0.353	0.367	94.9
		$\alpha_2 = 0.20$	0.190		-0.010	0.206	0.207	95.2
		$\sigma_v = 0.30$	0.250		-0.050	0.295	0.210	99.1
		$\tau = 0.0000$	0.003		0.003	0.087	0.087	95.3

Continued...Table 3.1

(K, n_i)	Conv.Prop.	Params	SM	SMP	Bias	ESE	SSE	Cov.Pr.
(100, 30)		$\beta_0 = 0.20$	0.207		0.007	0.052	0.066	94.7
		$\beta_1 = 0.40$	0.401		0.001	0.059	0.060	95.2
		$\beta_2 = 0.30$	0.297		-0.003	0.035	0.034	95.5
		$\sigma_u = 0.40$	0.392		-0.008	0.024	0.040	95.1
	0.913	$\alpha_0 = -1.50$	-1.500	0.501	0.000	0.182	0.178	93.8
		$\alpha_1 = 0.30$	0.303		0.003	0.244	0.239	94.9
		$\alpha_2 = 0.20$	0.190		-0.010	0.142	0.140	94.6
		$\sigma_v = 0.30$	0.292		-0.008	0.153	0.140	89.4
		$\tau = 0.0000$	0.004		0.004	0.060	0.058	94.0
(200,5)		$\beta_0 = 0.20$	0.202		0.002	0.081	0.080	95.1
		$\beta_1 = 0.40$	0.401		0.001	0.116	0.115	95.1
		$\beta_2 = 0.30$	0.297		-0.003	0.067	0.067	94.6
		$\sigma_u = 0.40$	0.390		-0.010	0.048	0.052	95.1
	0.900	$\alpha_0 = -1.50$	-1.506	0.506	-0.006	0.326	0.338	94.9
		$\alpha_1 = 0.30$	0.302		0.002	0.440	0.479	95.1
		$\alpha_2 = 0.20$	0.185		-0.015	0.256	0.267	94.4
		$\sigma_c = 0.30$	0.292		-0.008	0.431	0.308	96.4
		$\tau = 0.0000$	0.026		0.026	0.122	0.121	94.4
(200,15)		$\beta_0 = 0.20$	0.197		-0.003	0.049	0.052	94.6
		$\beta_1 = 0.40$	0.400		-0.000	0.061	0.062	95.1
		$\beta_2 = 0.30$	0.301		0.001	0.036	0.036	95.5
		$\sigma_u = 0.40$	0.394		-0.006	0.025	0.032	95.2
	0.932	$\alpha_0 = -1.50$	-1.505	0.502	-0.005	0.182	0.182	95.4
		$\alpha_1 = 0.30$	0.298		-0.002	0.248	0.251	95.7
		$\alpha_2 = 0.20$	0.197		-0.003	0.144	0.141	95.7
		$\sigma_c = 0.30$	0.276		-0.024	0.211	0.167	99.7
		$\tau = 0.0000$	0.006		0.006	0.062	0.064	94.8
(200,30)		$\beta_0 = 0.20$	0.206		0.006	0.038	0.050	94.8
		$\beta_1 = 0.40$	0.398		-0.002	0.042	0.042	95.1
		$\beta_2 = 0.30$	0.300		0.000	0.024	0.024	95.2
		$\sigma_u = 0.40$	0.387		-0.013	0.017	0.030	93.8
	0.922	$\alpha_0 = -1.50$	-1.496	0.502	0.004	0.128	0.127	95.0
		$\alpha_1 = 0.30$	0.301		0.001	0.173	0.169	95.4
		$\alpha_2 = 0.20$	0.202		0.002	0.100	0.098	95.8
		$\sigma_c = 0.30$	0.313		0.013	0.097	0.088	95.2
		$\tau = 0.0000$	0.008		0.008	0.042	0.042	94.7

Table 3.2: Simulated mean (SM), simulated mean of mixing proportion (SMP), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) for estimating Marginal Parameters: $(\boldsymbol{\beta}, \sigma_u)$; Component-1 Parameters $(\boldsymbol{\alpha}, \sigma_c)$ and τ for mixing proportion $\pi^* = 0.70$ from REMPois-Pois model for different number of clusters and various cluster sizes

(K, n_i)	Conv.Prop.	Params	SM	SMP	Bias	ESE	SSE	Cov.Pr.
(50,5)		$\beta_0 = 0.20$	0.206		0.006	0.183	0.194	94.7
		$\beta_1 = 0.40$	0.388		-0.012	0.259	0.278	94.1
		$\beta_2 = 0.30$	0.302		0.002	0.149	0.158	95.9
		$\sigma_u = 0.40$	0.371		-0.029	0.103	0.142	93.9
	0.884	$\alpha_0 = -1.50$	-1.507	0.704	-0.007	0.424	0.462	94.8
		$\alpha_1 = 0.30$	0.258		-0.042	0.621	0.715	94.5
		$\alpha_2 = 0.20$	0.184		-0.016	0.361	0.398	94.5
		$\sigma_c = 0.30$	0.277		-0.023	0.419	0.329	95.4
		$\tau = 0.8473$	0.868		0.020	0.184	0.195	94.8
(50,15)		$\beta_0 = 0.20$	0.212		0.012	0.124	0.131	95.8
		$\beta_1 = 0.40$	0.395		-0.005	0.162	0.166	94.6
		$\beta_2 = 0.30$	0.298		-0.002	0.093	0.094	95.2
		$\sigma_u = 0.40$	0.398		-0.002	0.068	0.098	94.6
	0.922	$\alpha_0 = -1.50$	-1.518	0.699	-0.018	0.291	0.291	94.1
		$\alpha_1 = 0.30$	0.297		-0.003	0.431	0.441	94.8
		$\alpha_2 = 0.20$	0.193		-0.007	0.247	0.240	95.0
		$\sigma_c = 0.30$	0.263		-0.037	0.291	0.218	98.4
		$\tau = 0.8473$	0.842		-0.005	0.125	0.129	96.0
(50,30)		$\beta_0 = 0.20$	0.211		0.011	0.074	0.105	95.3
		$\beta_1 = 0.40$	0.405		0.005	0.082	0.088	94.7
		$\beta_2 = 0.30$	0.300		0.000	0.047	0.047	94.7
		$\sigma_u = 0.40$	0.396		-0.004	0.034	0.058	95.6
	0.912	$\alpha_0 = -1.50$	-1.507	0.700	-0.007	0.169	0.171	95.0
		$\alpha_1 = 0.30$	0.303		0.003	0.235	0.239	95.4
		$\alpha_2 = 0.20$	0.201		0.001	0.136	0.136	94.8
		$\sigma_c = 0.30$	0.287		-0.013	0.119	0.110	95.1
		$\tau = 0.8473$	0.848		0.001	0.068	0.071	94.2
(100, 5)		$\beta_0 = 0.20$	0.205		0.005	0.126	0.133	95.3
		$\beta_1 = 0.40$	0.396		-0.004	0.175	0.186	95.0
		$\beta_2 = 0.30$	0.300		-0.000	0.102	0.106	95.0
		$\sigma_u = 0.40$	0.385		-0.015	0.072	0.086	94.8
	0.912	$\alpha_0 = -1.50$	-1.509	0.703	-0.009	0.292	0.292	95.3
		$\alpha_1 = 0.30$	0.310		0.010	0.424	0.432	94.7
		$\alpha_2 = 0.20$	0.181		-0.019	0.248	0.260	95.3
		$\sigma_v = 0.30$	0.257		-0.043	0.369	0.244	98.7
		$\tau = 0.8473$	0.864		0.016	0.133	0.140	94.7
(100, 15)		$\beta_0 = 0.20$	0.209		0.009	0.073	0.079	95.4
		$\beta_1 = 0.40$	0.397		-0.003	0.086	0.085	94.9
		$\beta_2 = 0.30$	0.300		0.000	0.050	0.051	94.7
		$\sigma_u = 0.40$	0.397		-0.003	0.036	0.052	94.7
	0.918	$\alpha_0 = -1.50$	-1.494	0.701	0.006	0.167	0.174	94.4
		$\alpha_1 = 0.30$	0.285		-0.015	0.234	0.241	95.2
		$\alpha_2 = 0.20$	0.203		0.003	0.137	0.138	94.6
		$\sigma_v = 0.30$	0.275		-0.025	0.156	0.142	87.9
		$\tau = 0.8473$	0.853		0.006	0.070	0.071	95.1

Continued...Table 3.2

(K, n_i)	Conv.Prop.	Params	SM	SMP	Bias	ESE	SSE	Cov.Pr.
(100, 30)		$\beta_0 = 0.20$	0.211		0.011	0.053	0.074	94.8
		$\beta_1 = 0.40$	0.397		-0.003	0.057	0.057	94.5
		$\beta_2 = 0.30$	0.299		-0.001	0.033	0.034	94.2
		$\sigma_u = 0.40$	0.390		-0.010	0.024	0.045	95.0
	0.929	$\alpha_0 = -1.50$	-1.500	0.701	0.000	0.118	0.118	95.2
		$\alpha_1 = 0.30$	0.293		-0.007	0.164	0.166	94.7
		$\alpha_2 = 0.20$	0.203		0.003	0.096	0.098	95.2
		$\sigma_v = 0.30$	0.297		-0.003	0.079	0.071	95.7
		$\tau = 0.8473$	0.851		0.004	0.048	0.047	95.2
(200,5)		$\beta_0 = 0.20$	0.198		-0.002	0.090	0.092	94.2
		$\beta_1 = 0.40$	0.405		0.005	0.122	0.127	95.7
		$\beta_2 = 0.30$	0.303		0.003	0.071	0.075	95.1
		$\sigma_u = 0.40$	0.391		-0.009	0.049	0.055	94.2
	0.933	$\alpha_0 = -1.50$	-1.509	0.701	-0.009	0.208	0.206	95.1
		$\alpha_1 = 0.30$	0.303		0.003	0.297	0.302	95.2
		$\alpha_2 = 0.20$	0.193		-0.007	0.172	0.178	94.6
		$\sigma_c = 0.30$	0.272		-0.028	0.276	0.203	99.2
		$\tau = 0.8473$	0.852		0.005	0.093	0.094	95.2
(200,15)		$\beta_0 = 0.20$	0.204		0.004	0.052	0.056	94.5
		$\beta_1 = 0.40$	0.398		-0.002	0.061	0.060	94.2
		$\beta_2 = 0.30$	0.299		-0.001	0.036	0.036	94.5
		$\sigma_u = 0.40$	0.394		-0.006	0.026	0.035	94.6
	0.926	$\alpha_0 = -1.50$	-1.498	0.701	0.002	0.117	0.112	95.8
		$\alpha_1 = 0.30$	0.300		-0.000	0.166	0.166	95.6
		$\alpha_2 = 0.20$	0.198		-0.002	0.097	0.098	95.9
		$\sigma_c = 0.30$	0.300		0.000	0.097	0.085	96.4
		$\tau = 0.8473$	0.853		0.005	0.050	0.051	94.4
(200,30)		$\beta_0 = 0.20$	0.209		0.009	0.038	0.052	95.0
		$\beta_1 = 0.40$	0.400		0.000	0.041	0.043	95.3
		$\beta_2 = 0.30$	0.299		-0.001	0.024	0.024	95.1
		$\sigma_u = 0.40$	0.383		-0.017	0.017	0.031	92.8
	0.960	$\alpha_0 = -1.50$	-1.501	0.701	-0.001	0.084	0.086	94.5
		$\alpha_1 = 0.30$	0.307		0.007	0.117	0.120	94.9
		$\alpha_2 = 0.20$	0.200		0.000	0.068	0.069	95.7
		$\sigma_c = 0.30$	0.303		0.003	0.054	0.051	94.6
		$\tau = 0.8473$	0.854		0.007	0.034	0.034	93.5

The amount of bias of τ was 0.082, 0.039 and 0.005 when $n_i = 5$, 15 and 30, respectively for $K = 50$. These amounts were 0.052, 0.003 and 0.004 for $n_i = 5$, 15 and 30, respectively when $K = 100$; these figures were 0.026, 0.006 and 0.008 for $n_i = 5$, 15 and 30, respectively if $K = 200$.

The convergence rate increases with increasing the cluster size. For example, the convergence rate of the ML estimation technique was 0.789, 0.846 and 0.920 when $n_i = 5$, 15 and 30, respectively for $K = 50$ whereas, these amount were 0.859, 0.938 and 0.913 for $n_i = 5$, 15 and 30, respectively when $K = 100$; these were 0.900, 0.932 and 0.922 for $n_i = 5$, 15 and 30, respectively if $K = 200$. It was found that the coverage probabilities always lied in the nominal level of confidence for all the setups.

Extensive simulation studies have also been conducted for mixing proportion, $\pi^* = 0.70$ and 0.90. The results of simulation studies for $\pi^* = 0.70$ and 0.90 were presented in Table 3.2 and Table 3.3, respectively. The results obtained from Table 3.2, and Table 3.3 are very similar with the results obtained from Table 3.1 but, the convergence rates went up for $\pi^* = 0.70$ and 0.90. The convergence rates were lied from 84% to 97% depending on the simulation setup.

Table 3.3: Simulated mean (SM), simulated mean of mixing proportion (SMP), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) for estimating Marginal Parameters: $(\boldsymbol{\beta}, \sigma_u)$; Component-1 Parameters $(\boldsymbol{\alpha}, \sigma_c)$ and τ for mixing proportion $\pi^* = 0.90$ from REMPois-Pois model for different number of clusters and various cluster sizes

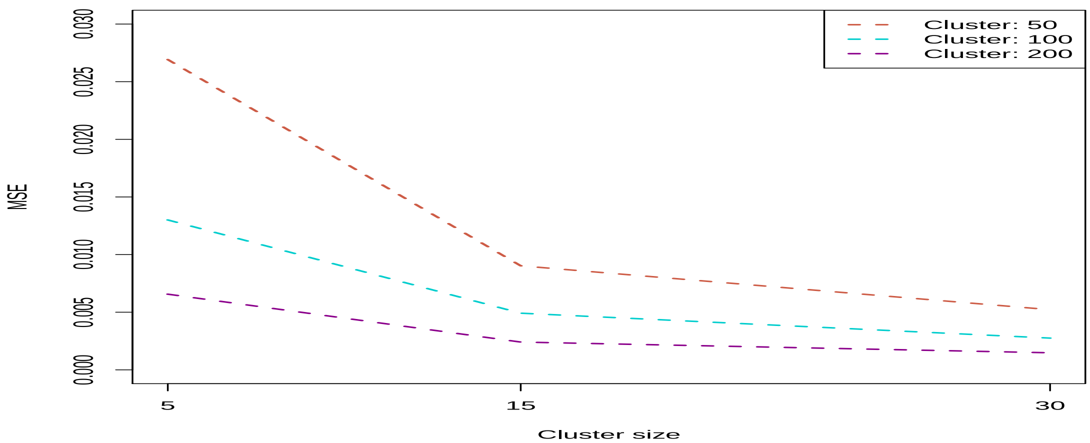
(K, n_i)	Conv.Prop.	Params	SM	SMP	Bias	ESE	SSE	Cov.Pr.
(50,5)		$\beta_0 = 0.20$	0.156		-0.044	0.233	0.367	95.1
		$\beta_1 = 0.40$	0.415		0.015	0.279	0.458	95.7
		$\beta_2 = 0.30$	0.308		0.008	0.158	0.244	94.7
		$\sigma_u = 0.40$	0.471		0.071	0.097	0.323	95.7
	0.845	$\alpha_0 = -1.50$	-1.521	0.896	-0.021	0.301	0.306	95.7
		$\alpha_1 = 0.30$	0.280		-0.020	0.461	0.483	94.2
		$\alpha_2 = 0.20$	0.210		0.010	0.267	0.271	94.7
		$\sigma_c = 0.30$	0.217		-0.083	0.312	0.243	98.8
		$\tau = 2.1972$	2.150		-0.047	0.215	0.229	94.4
(50,15)		$\beta_0 = 0.20$	0.212		0.012	0.127	0.157	94.7
		$\beta_1 = 0.40$	0.396		-0.004	0.126	0.145	95.1
		$\beta_2 = 0.30$	0.300		-0.000	0.073	0.084	95.0
		$\sigma_u = 0.40$	0.408		0.008	0.049	0.098	94.6
	0.913	$\alpha_0 = -1.50$	-1.501	0.900	-0.001	0.166	0.166	94.9
		$\alpha_1 = 0.30$	0.300		-0.000	0.249	0.252	95.7
		$\alpha_2 = 0.20$	0.202		0.002	0.146	0.147	95.0
		$\sigma_c = 0.30$	0.277		-0.023	0.130	0.123	91.5
		$\tau = 2.1972$	2.198		0.000	0.123	0.119	95.6
(50,30)		$\beta_0 = 0.20$	0.215		0.015	0.088	0.125	94.2
		$\beta_1 = 0.40$	0.398		-0.002	0.077	0.084	94.4
		$\beta_2 = 0.30$	0.303		0.003	0.045	0.049	95.1
		$\sigma_u = 0.40$	0.392		-0.008	0.031	0.072	95.6
	0.926	$\alpha_0 = -1.50$	-1.505	0.900	-0.005	0.124	0.126	94.5
		$\alpha_1 = 0.30$	0.303		0.003	0.174	0.180	95.6
		$\alpha_2 = 0.20$	0.202		0.002	0.101	0.102	95.2
		$\sigma_c = 0.30$	0.297		-0.003	0.075	0.071	94.3
		$\tau = 2.1972$	2.199		0.002	0.087	0.085	95.5
(100, 5)		$\beta_0 = 0.20$	0.194		-0.006	0.168	0.195	94.4
		$\beta_1 = 0.40$	0.408		0.008	0.194	0.244	95.3
		$\beta_2 = 0.30$	0.306		0.006	0.112	0.135	94.0
		$\sigma_u = 0.40$	0.407		0.007	0.064	0.114	95.9
	0.903	$\alpha_0 = -1.50$	-1.509	0.900	-0.009	0.206	0.208	94.8
		$\alpha_1 = 0.30$	0.300		-0.000	0.310	0.314	94.2
		$\alpha_2 = 0.20$	0.195		-0.005	0.181	0.180	94.9
		$\sigma_v = 0.30$	0.244		-0.056	0.254	0.182	99.6
		$\tau = 2.1972$	2.194		-0.003	0.152	0.150	95.2
(100, 15)		$\beta_0 = 0.20$	0.207		0.007	0.092	0.113	94.7
		$\beta_1 = 0.40$	0.401		0.001	0.088	0.102	94.5
		$\beta_2 = 0.30$	0.301		0.001	0.051	0.055	95.9
		$\sigma_u = 0.40$	0.397		-0.003	0.033	0.062	96.5
	0.920	$\alpha_0 = -1.50$	-1.500	0.900	-0.000	0.121	0.121	94.6
		$\alpha_1 = 0.30$	0.300		-0.000	0.174	0.174	95.1
		$\alpha_2 = 0.20$	0.206		0.006	0.102	0.096	94.2
		$\sigma_v = 0.30$	0.289		-0.011	0.086	0.078	96.2
		$\tau = 2.1972$	2.196		-0.001	0.087	0.085	95.7

Continued...Table 3.3

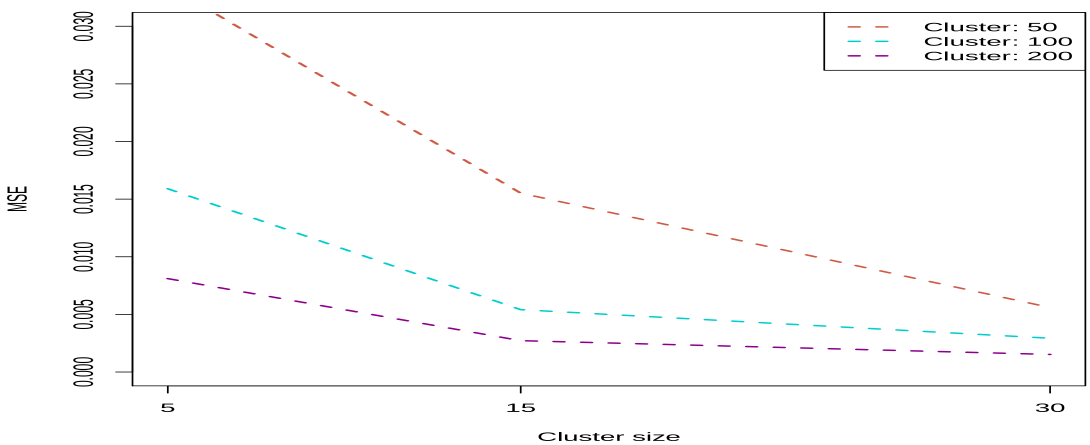
(K, n_i)	Conv.Prop.	Params	SM	SMP	Bias	ESE	SSE	Cov.Pr.
(100, 30)		$\beta_0 = 0.20$	0.212		0.012	0.063	0.093	94.6
		$\beta_1 = 0.40$	0.401		0.001	0.055	0.058	95.3
		$\beta_2 = 0.30$	0.300		0.000	0.032	0.036	94.4
		$\sigma_u = 0.40$	0.378		-0.022	0.022	0.049	94.5
	0.970	$\alpha_0 = -1.50$	-1.496	0.901	0.004	0.087	0.086	95.8
		$\alpha_1 = 0.30$	0.300		-0.000	0.122	0.123	95.4
		$\alpha_2 = 0.20$	0.197		-0.003	0.071	0.071	94.0
		$\sigma_v = 0.30$	0.297		-0.003	0.052	0.052	95.4
		$\tau = 2.1972$	2.204		0.007	0.062	0.060	94.9
(200,5)		$\beta_0 = 0.20$	0.204		0.004	0.119	0.128	95.0
		$\beta_1 = 0.40$	0.401		0.001	0.136	0.154	94.4
		$\beta_2 = 0.30$	0.300		0.000	0.079	0.089	94.6
		$\sigma_u = 0.40$	0.396		-0.004	0.043	0.061	95.2
	0.929	$\alpha_0 = -1.50$	-1.504	0.900	-0.004	0.146	0.145	94.2
		$\alpha_1 = 0.30$	0.290		-0.010	0.216	0.217	94.0
		$\alpha_2 = 0.20$	0.205		0.005	0.126	0.134	94.3
		$\sigma_c = 0.30$	0.255		-0.045	0.182	0.152	85.0
		$\tau = 2.1972$	2.195		-0.002	0.107	0.108	96.1
(200,15)		$\beta_0 = 0.20$	0.204		0.004	0.065	0.076	95.0
		$\beta_1 = 0.40$	0.399		-0.001	0.062	0.065	95.3
		$\beta_2 = 0.30$	0.298		-0.002	0.036	0.039	94.8
		$\sigma_u = 0.40$	0.384		-0.016	0.023	0.037	93.1
	0.956	$\alpha_0 = -1.50$	-1.494	0.901	0.006	0.085	0.083	93.8
		$\alpha_1 = 0.30$	0.294		-0.006	0.123	0.122	94.7
		$\alpha_2 = 0.20$	0.198		-0.002	0.072	0.071	95.4
		$\sigma_c = 0.30$	0.298		-0.002	0.058	0.056	94.8
		$\tau = 2.1972$	2.204		0.007	0.062	0.061	95.2
(200,30)		$\beta_0 = 0.20$	0.205		0.005	0.046	0.066	94.7
		$\beta_1 = 0.40$	0.401		0.001	0.039	0.041	94.8
		$\beta_2 = 0.30$	0.299		-0.001	0.023	0.025	96.6
		$\sigma_u = 0.40$	0.370		-0.030	0.015	0.032	86.1
	0.965	$\alpha_0 = -1.50$	-1.497	0.901	0.003	0.062	0.058	94.6
		$\alpha_1 = 0.30$	0.298		-0.002	0.087	0.085	95.5
		$\alpha_2 = 0.20$	0.200		-0.000	0.050	0.051	95.1
		$\sigma_c = 0.30$	0.303		0.003	0.036	0.035	94.7
		$\tau = 2.1972$	2.207		0.010	0.044	0.045	94.0

The MSE of $\beta_0, \beta_1, \beta_2, \sigma_u$ were shown in Figures 3.1-3.4, respectively. In each figure, part (a), part (b) and part (c) were drawn for $\pi^* = 0.50, 0.70$ and 0.90, respectively. From these figures, it is depicted that the MSEs were very close to zero for all the indicated situations when cluster size is considered as 15 or 30. Also, the MSEs decreases for increasing cluster size and/or number of clusters.

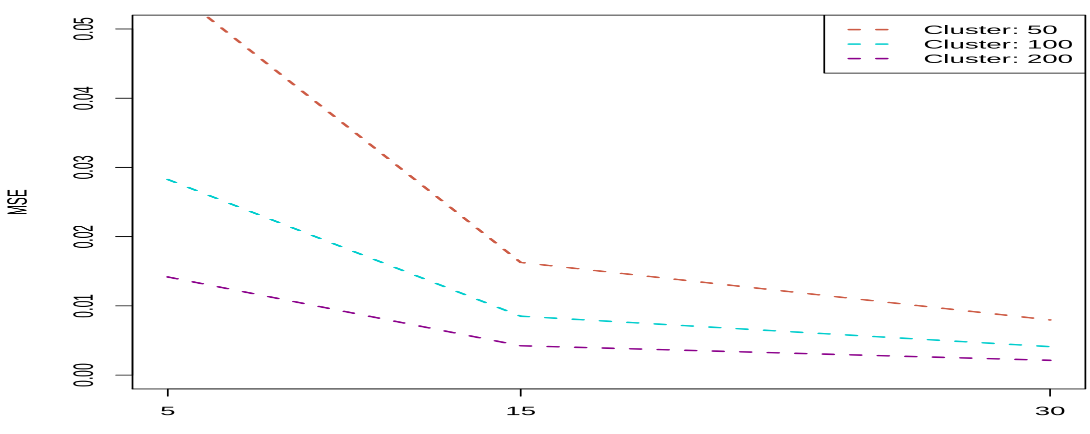
The intra-cluster correlation coefficient (ICC) and its estimated value were computed using Eq.(3.16) by employing the values of the true parameters and their estimates, respectively. The ICC and estimates of ICC was given in the Appendix B (Section B.4) for different number of clusters and various cluster sizes and for varying probability of mixture in Table-B.1. Using Eq.(3.14)-Eq.(3.16), it can be easily observed that the ICC decreases as the unconditional variance of Y_{ij} increases, and the unconditional variance of Y_{ij} increases with increasing the mixing probability. The true values of ICC were range from 0.10747 to 0.10873 and its estimates were from 0.09832 to 0.10785 when we considered $\pi^* = 0.50$; the true ICCs varied from 0.06224 to 0.06258 and its estimates were from 0.05792 to 0.06465 when $\pi^* = 0.70$; and the values of ICC went from 0.02006 to 0.02008 and its estimates were from 0.01743 to 0.03697 when $\pi^* = 0.90$ for different number of clusters and various cluster sizes. Although the ICC can varies 0 to 1 theoretically, the ICCs can be observed as small value such as 0.01 or 0.02 for discrete clustered data in most of the human studies (Peerawaranun et al., 2019; Killip et al., 2004; Murray and Short, 1997).



(a) Cluster size vs. MSE for $\pi^* = 0.50$

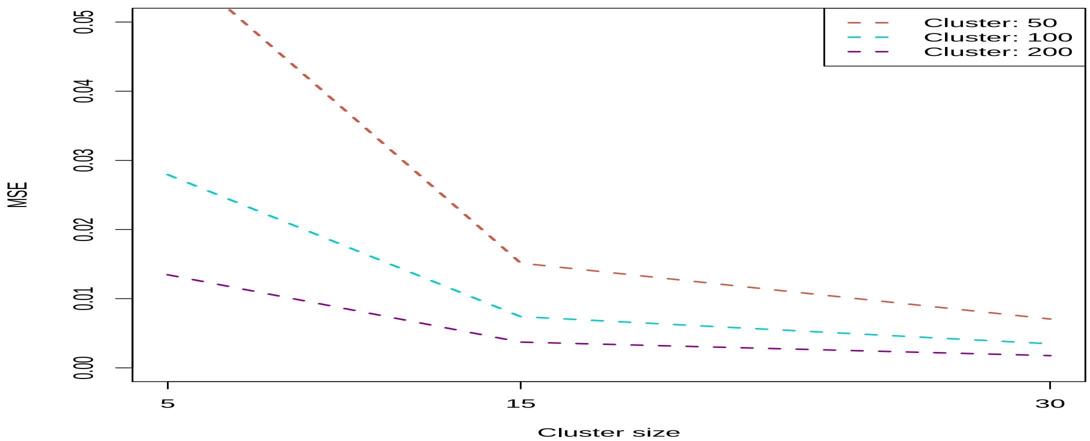


(b) Cluster size vs. MSE for $\pi^* = 0.70$

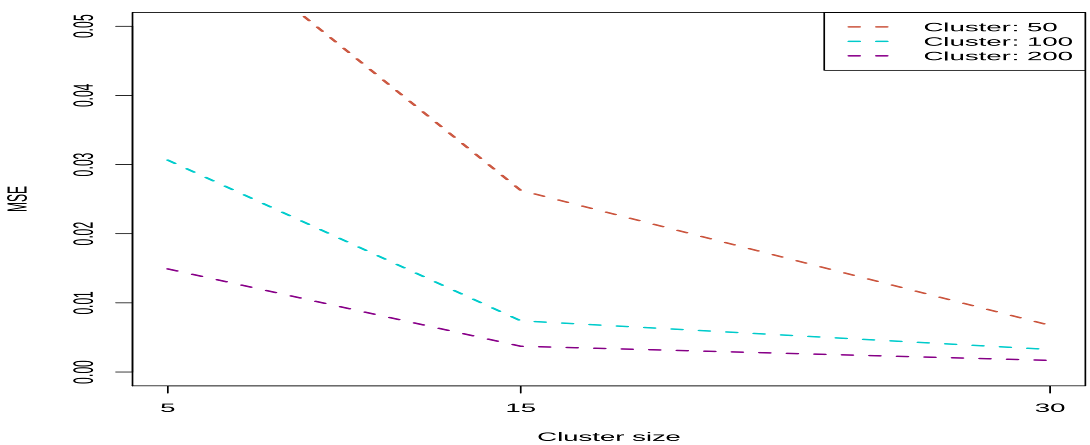


(c) Cluster size vs. MSE for $\pi^* = 0.90$

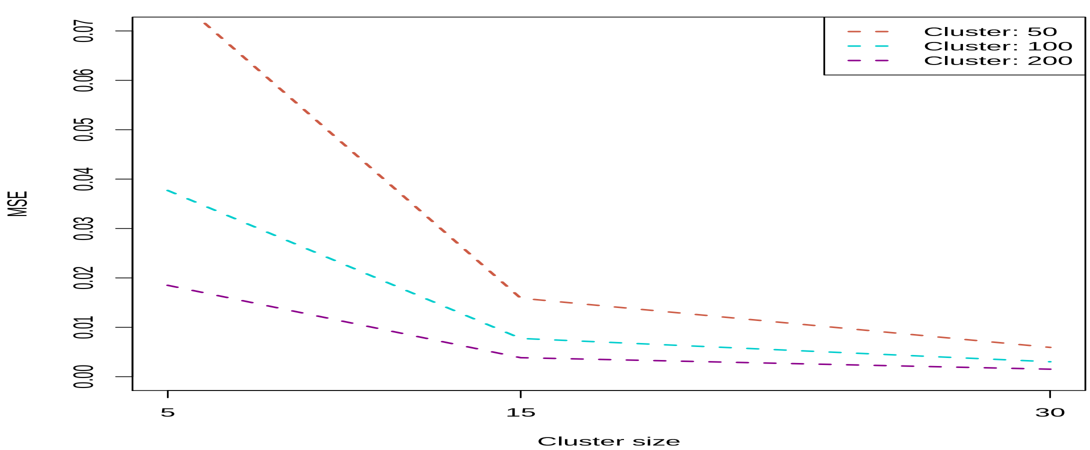
Figure 3.1: A comparison of MSE from REMPois-Pois models with different mixing probability and varying cluster sizes for the regression parameter $\beta_0 = 0.20$



(a) Cluster size vs. MSE for $\pi^* = 0.50$

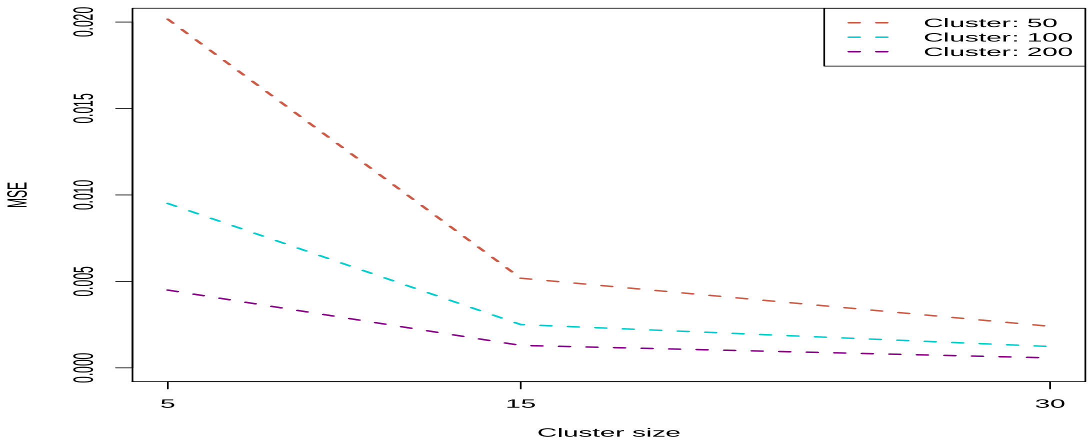


(b) Cluster size vs. MSE for $\pi^* = 0.70$

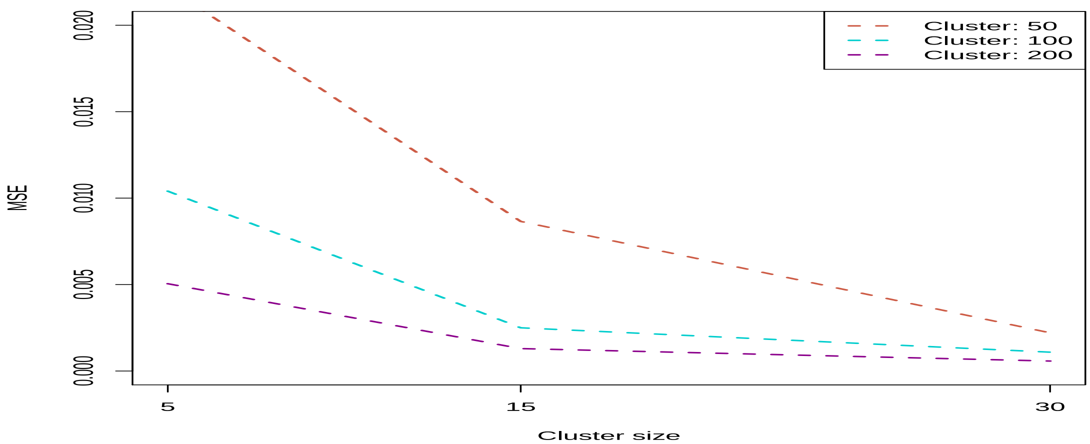


(c) Cluster size vs. MSE for $\pi^* = 0.90$

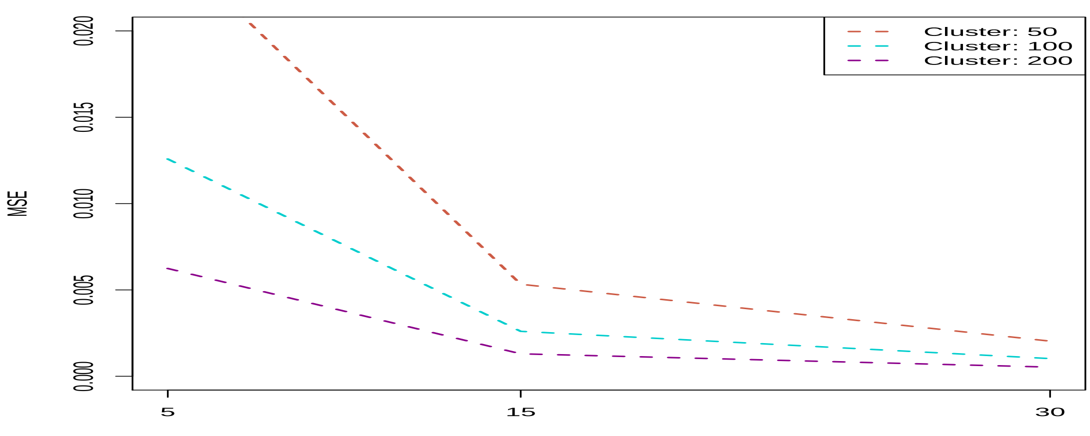
Figure 3.2: A comparison of MSE from REMPois-Pois models with different mixing probability and varying cluster sizes for the regression parameter $\beta_1 = 0.40$



(a) Cluster size vs. MSE for $\pi^* = 0.50$

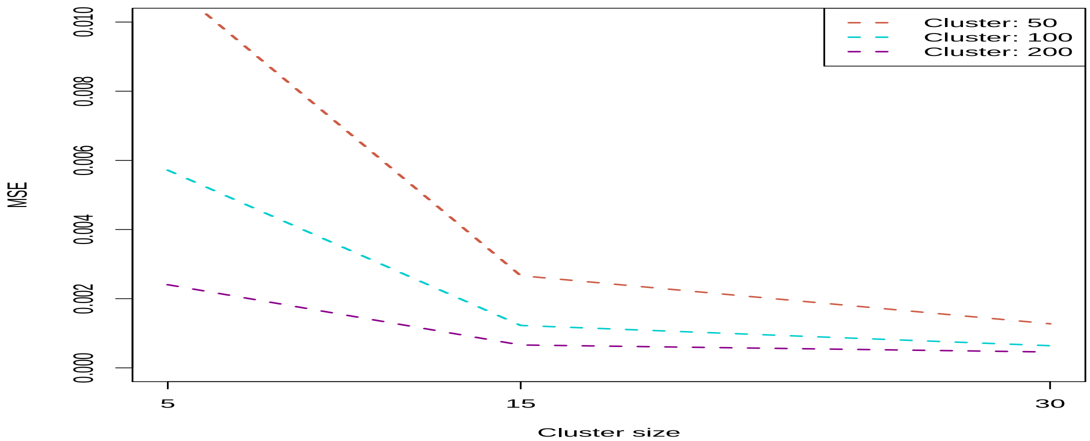


(b) Cluster size vs. MSE for $\pi^* = 0.70$

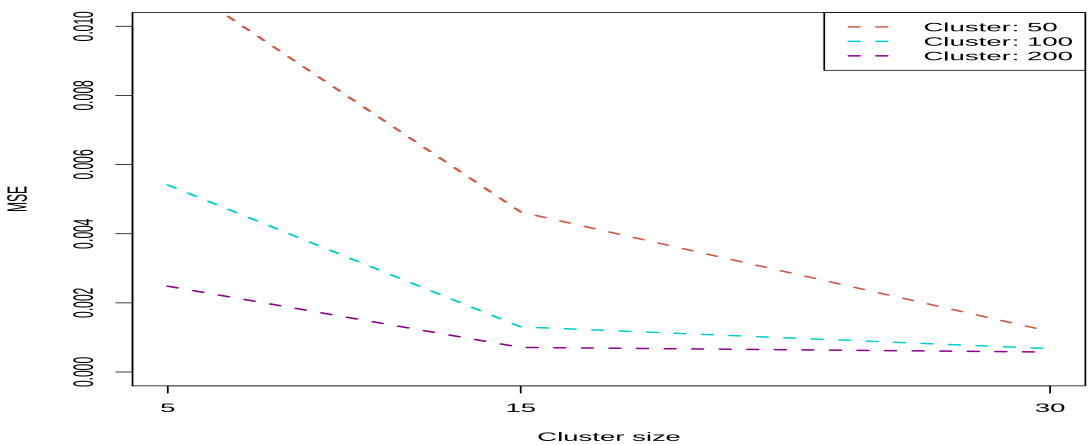


(c) Cluster size vs. MSE for $\pi^* = 0.90$

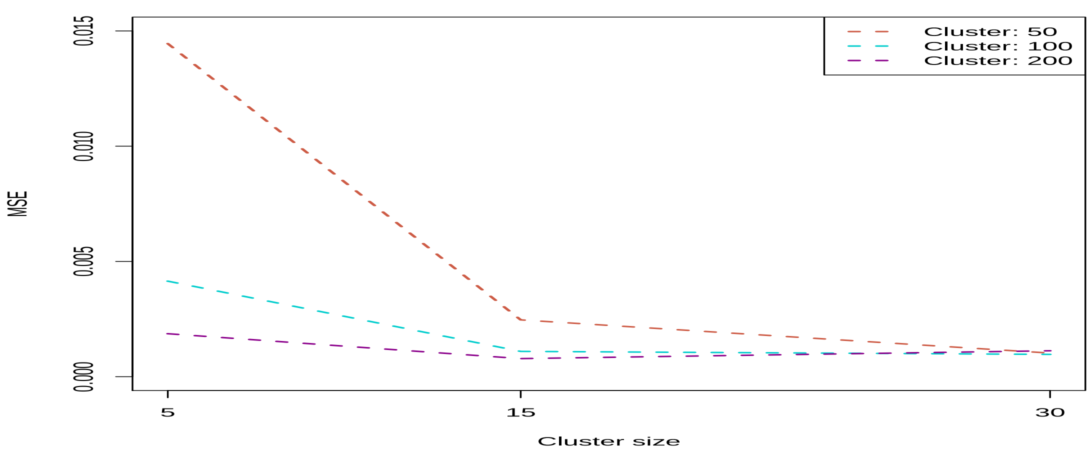
Figure 3.3: A comparison of MSE from REMPois-Pois models with different mixing probability and varying cluster sizes for the regression parameter $\beta_2 = 0.30$



(a) Cluster size vs. MSE for $\pi^* = 0.50$



(b) Cluster size vs. MSE for $\pi^* = 0.70$



(c) Cluster size vs. MSE for $\pi^* = 0.90$

Figure 3.4: A comparison of MSE from REMPois-Pois models with different mixing probability and varying cluster sizes for the regression parameter $\sigma_u = 0.40$

3.2.2 The situation when $\rho^* > 0$

A simulation study was also conducted for evaluating the performance of the proposed model by considering bivariate random effects for fixed number of 200 clusters each with size 10. In this case, a low ($\rho^* = 0.30$) degree and a high ($\rho^* = 0.70$) degree of positive correlation between the random effects for marginal means and random effects for component-1 means had been considered. These random effects had been utilized in generating two-component Poisson count clustered data only by using $\tau = 0.8473$ ($\pi^* = 0.70$). At the step-1 of simulation study, K bivariate random variables, $(U_1, C_1), \cdots, (U_K, C_K)$ from $N_2(0, \Sigma)$ had been generated and each paired variable was then repeated n_i times ($i = 1, \cdots, K$) to make up the n values of U and C. Since the values of all other parameters were considered same as the simulation study for $\rho^* = 0$, step-2 to step-6 of the simulation study for $\rho^* = 0$ were followed. Results of the simulation studies were presented in Table 3.4, which also reports the SM, SMP, Bias, ESE, SSE, Cov.Pr. and Conv.Prop. computed from 1000 replications. From Table 3.4, it was observed that the performance of the proposed model with bivariate random effects was similar as the performance of the proposed model with independent random effects. Also, the true values of ICC and its estimates were found similar as those values computed for $\rho^* = 0$ and for $\pi^* = 0.70$. However, the Conv.Prop. decreases while considering bivariate random effects.

Table 3.4: Results from simulation study for estimating marginal parameters: $(\boldsymbol{\beta}, \sigma_u)$; component-1 parameters $(\boldsymbol{\alpha}, \sigma_c)$ from REMPois-Pois model by considering correlated $\boldsymbol{u}$ and $\boldsymbol{c}$ (random effects for marginal and subpopulation) with mixing proportion $\pi^* = 0.70$ and cluster size 10 for 200 clusters

Conv.Prop.	Params	SM	SMP	Bias	ESE	SSE	Cov.Pr.
0.732	$\beta_0 = 0.20$	0.199	0.701	-0.001	0.064	0.061	94.7
	$\beta_1 = 0.40$	0.400		0.000	0.079	0.080	94.7
	$\beta_2 = 0.30$	0.304		0.004	0.046	0.046	94.8
	$\sigma_u = 0.40$	0.398		-0.002	0.037	0.038	95.5
	$\alpha_0 = -1.50$	-1.503		-0.003	0.146	0.142	95.5
	$\alpha_1 = 0.30$	0.307		0.007	0.204	0.210	95.2
	$\alpha_2 = 0.20$	0.205		0.005	0.119	0.120	95.6
	$\sigma_c = 0.30$	0.272		-0.028	0.119	0.133	94.4
	$\rho^*_{unres} = 0.3568$	0.468		0.112	0.215	0.284	91.4
	$\tau = 0.8473$	0.852		0.004	0.062	0.060	94.9
	$\rho^* = 0.30$	0.367		0.067			
0.767	$\beta_0 = 0.20$	0.202	0.700	0.002	0.064	0.063	95.0
	$\beta_1 = 0.40$	0.402		0.002	0.079	0.079	94.1
	$\beta_2 = 0.30$	0.300		0.000	0.046	0.043	94.5
	$\sigma_u = 0.40$	0.401		0.001	0.037	0.036	94.3
	$\alpha_0 = -1.50$	-1.510		-0.010	0.148	0.147	94.1
	$\alpha_1 = 0.30$	0.303		0.003	0.204	0.209	94.5
	$\alpha_2 = 0.20$	0.204		0.004	0.119	0.114	95.3
	$\sigma_c = 0.30$	0.323		0.023	0.100	0.097	95.0
	$\rho^*_{unres} = 1.204$	1.153		-0.051	0.320	0.385	96.7
	$\tau = 0.8473$	0.849		0.002	0.061	0.064	94.9
	$\rho^* = 0.70$	0.660		-0.040			

3.3 Illustration

To illustrate the application of REMPois-Pois model, a nationwide representative data extracted from the 2014 Bangladesh Demographic and Health Survey (BDHS) have been utilized. The data and variables were described in the Section 2.5.

3.3.1 Results

The mean, standard deviation, minimum and maximum of the number of ANC visits for the data under consideration were 2.78, 2.56, 0 and 20, respectively. The distribution of

Table 3.5: The estimated parameters and other statistics for REMPois-Pois models for analyzing the number of ANC visits during a pregnancy period, BDHS 2014

Variable name	Estimates	SE	p-value	IRR
Marginalized Poisson-Poisson model:				
Intercept	0.340	0.066	<0.001	
Area of residence				
Rural (ref)				
Urban	0.204	0.038	<0.001	1.226
Level of education				
No education (ref)				
Primary	0.256	0.058	<0.001	1.291
Secondary	0.458	0.057	<0.001	1.581
Higher	0.659	0.064	<0.001	1.933
Media exposure				
Not exposure (ref)				
Exposure	0.174	0.032	<0.001	1.190
Maternal age (years)				
<20	-0.079	0.033	0.016	0.924
20-29 (ref)				
≥30	0.060	0.042	0.153	1.062
Gap between husband age and wife age (years)				
Non-positive	0.186	0.110	0.092	1.205
1-5 (ref)				
6-10	0.002	0.029	0.953	1.002
>10	-0.010	0.032	0.766	0.990
Number of reasons wife beating justified				
Not at all (ref)				
1-2	-0.019	0.033	0.561	0.981
3-5	-0.055	0.049	0.264	0.946
Wealth index				
Poor (ref)				
Middle	0.133	0.039	0.001	1.142
Rich	0.328	0.044	<0.001	1.388
Birth Order				
1 (ref)				
2	-0.065	0.032	0.044	0.937
3	-0.132	0.045	0.003	0.876
≥4	-0.276	0.059	<0.001	0.759
Random effect: σ_u	0.297	0.016	<0.001	
Corr(marginal,subpopulation): $\rho^*_{u,c}$	0.844			
Intra-cluster correlation: ρ	0.124			
AIC	17737			

Continued...Table 3.5

Variable name	Estimates	SE	p-value	IRR
Model for Component-1:				
Intercept	-0.790	0.198	<0.001	
Area of residence				
Rural (ref)				
Urban	0.259	0.071	<0.001	-
Level of education				
No education (ref)				
Primary	0.420	0.161	0.009	-
Secondary	0.904	0.165	<0.001	-
Higher	1.151	0.178	<0.001	-
Media exposure				
Not exposure (ref)				
Exposure	0.290	0.069	<0.001	-
Maternal age (years)				
<20	-0.214	0.069	0.002	-
20-29 (ref)				
≥30	0.232	0.084	0.006	-
Gap between husband age and wife age (years)				
Non-positive	0.232	0.084	0.006	-
1-5 (ref)				
6-10	0.077	0.062	0.213	-
>10	0.149	0.070	0.032	-
Number of reasons wife beating justified				
Not at all (ref)				
1-2	-0.051	0.065	0.433	-
3-5	-0.239	0.104	0.022	-
Wealth index				
Poor (ref)				
Middle	0.402	0.082	<0.001	-
Rich	0.699	0.095	<0.001	-
Birth Order				
1 (ref)				
2	-0.240	0.071	0.001	-
3	-0.335	0.097	0.001	-
≥4	-0.693	0.128	<0.001	-
Random effect:σ_c	0.431	0.018	<0.001	
Mixing Proportion: π^*	0.582			

the number of ANC visits is presented in Figure 2.4 which indicates the zero-inflated model would be suited for the data. At first a null model had been fitted to investigate the applicability of the proposed REMPois-Pois model in the data under consideration. From the null model, it was observed that the frequency of ANC visits in Bangladesh arises from mixture of two unobserved populations with mixing proportion $\pi^* = 0.66$; the random effects had significant influence on both the marginal model as well as subpopulations; and the random effects possess high positive correlation (0.85). This findings motivated us to fit a REMPois-Pois model for analyzing data under consideration. For the full model, the AIC of fitted REMPois-Pois model was found as 17737 while it was 17757 for random effects negative binomial model and 18372 for random effects Poisson model. The main aim of this illustration is to find some important determinants of the outcome variable under study as well as to estimate the intra-cluster correlation (ρ).

The results of the fitted REMPois-Pois model are presented in Table 3.5. From Table 3.5, it was found that the covariates *area of residence, level of education, media exposure, maternal age, gap between husband age and wife age (years), wealth index, birth order* are the factors had significant influence on the *number of ANC visits* during the pregnancy period of a woman; the intra-cluster correlation under the fitted model was $\hat{\rho} = 0.124$; and the correlation between the random effects for marginal model and subpopulation was ($\hat{\rho}^* = 0.844$). It was observed that, the estimated incidence rate of ANC visits was 22.6% higher for urban women as compared to rural women (p-value<0.001). The IRR of ANC visits were 1.291, 1.581, 1.933 for mothers with education level 'primary', 'secondary' and 'higher' respectively to mothers with education level 'no education'. All three categories of 'level of education' were statistically significant with p-value<0.001. The IRR of ANC visits was 1.19 for mothers with exposed media to unexposed media (p-value<0.001). The rate of ANC visits was 7.6% lower for mothers who gave their index birth below 20 years of age (p-value<0.05) than those who gave birth during age 20-29 years. However, the rate was statistically insignificant for mothers who gave birth at age ≥ 30 years than those who gave birth during age 20-29 years. The incidence rate of ANC visits were 20.5% higher for women whose husband were not older than them compared to the women whose husband were 1-5 years older than them . However, the rate was statistically insignificant for other categories of 'gap between husband age and wife age (years)'. The rates of ANC visits were statistically insignificant among all categories of 'number of reasons wife beating justified'. The rate of ANC visit was 14.2% higher for the mothers from middle-wealth households (p-value<0.01) than the poor households, and 38.8% higher for the mothers from rich households than the poor households (p-value<0.001). The rate of ANC visits of mothers during their pregnancy were 6.3%, 12.4% and 24.1% less likely for the second birth, third birth and forth or upper order birth respectively compared to the first birth. It was also experienced that the marginal mean parameters were estimated from mixture of two latent subpopulations with the proportions 0.58 and 0.42; the random effects had significant influence on both the marginal model ($\hat{\sigma}_u = 0.297$, p-value <0.001) as well as subpopulations ($\hat{\sigma}_c = 0.431$, p-value <0.001) ; and the random effects possess high positive correlation (0.84) after adjusting the covariates.

3.4 Conclusion

Like cross-sectional setup, it is impossible to obtain the inference regarding the exposure effects on the marginal mean (marginalized over subpopulations) from latent class regression model for clustered data. Therefore, the marginal inference from the existing Poisson-Poisson mixture model with random effects is hardly possible. In this chapter, we have developed a marginalized two components Poisson mixture model with random effects for clustered count data. The parameters of the proposed model have been estimated by using the ML method. However, this estimation technique requires evaluating the integrals while maximizing the log-likelihood function, which results in an intractable form of the marginal distribution of responses within a cluster. To address the difficulty, we have utilized the Gauss–Hermite quadrature method to approximate the integral for the random effects.

The proposed method enables us to directly model the marginal cluster specific mean and it controls latent class-based unexplained heterogeneity. Therefore, the proposed REMPois-Pois model provides inference of parameters that allows marginal interpretation rater than latent class interpretation of the parameters along with the inference of random effects parameters as well as the intra-cluster correlation coefficient.

The performance of the proposed model has been examined by conducting extensive simulation studies. It was observed from the simulation studies that the REMPois-Pois model provides estimates with low bias and low MSE when the number of clusters as well as the cluster size were considered large.

In this study, a nationally representative data set from Bangladesh has been analyzed using the proposed model as an illustration. It was found that the sampled data have been arisen from two latent subpopulations using mixing probability of 0.582. The potential predictors for the frequency of ANC visits were determined by fitting the proposed model. It was observed that the intraclass correlation between the random effects of the marginalized model and the component-1 model was high.

Chapter 4

Marginalized Mixture Models: Zero-Inflated Repeated Measures Count Data

Repeated measures data, known as longitudinal data, usually comprise of responses obtained from each individual repeatedly at multiple occasions. Since the observations are obtained from same individual, they are likely to exhibit positive correlation in longitudinal setup (Fitzmaurice et al., 2012). The method of analyzing such data is usually complex due to the relationship among the observed values of the response variable. However, various analytical methods have been developed by taking the correlation or covariance into account for analyzing discrete longitudinal data (Liang and Zeger, 1986; Fitzmaurice and Laird, 1993; Hall and Severini, 1998; Zhao and Prentice, 1990; Prentice and Zhao, 1991; Zhao et al., 1992; Sutradhar and Das, 1999).

The maximum likelihood (ML) estimation of regression parameters of the model fo the repeated count responses had been found difficult. This is because the joint probability distribution of repeated count responses has complicated functional form. To solve the difficulty, Liang and Zeger (1986) proposed a 'working' correlation structure based generalized estimating equation (GEE) to obtain the consistent estimates of regression parameters by using generalized linear model (GLM) framework. In this approach, the 'working' correlation parameters are estimated by 'method of moments' and these estimates are then use in the GEE for the regression parameters. However, the method proposed by Liang and Zeger (1986) may lead to a complete breakdown in the estimation method because of uncertainty of the definition of 'working' correlation (Crowder, 1995). Later, Fitzmaurice and Laird (1993); Hall and Severini (1998) have developed estimating equation for finding estimates of regression parameters and parameters for 'working' correlation simultaneously. Additionally, Zhao and Prentice (1990); Prentice and Zhao (1991); Zhao et al. (1992) have developed an extended GEE method for joint estimation of the regression parameters and the true correlation parameters. Although 'working' correlation based GEE estimators (Liang and Zeger, 1986) were developed with a view to gaining efficiency, Sutradhar and Das (1999); Sutradhar and Kumar (2001) have argued that the efficiency of these estimators are rather less than the efficiency of 'working' independence based estimators in many situations. In order to gaining efficiency of the estimators, Sutradhar and Das (1999) proposed a generalized quasi-likelihood (GQL) approach for the estimation of the regression parameters of the longitudinal model with true correlation structure and the correlation parameters are then estimated by using the method of moments. The estimating equations of the GQL appraoch use a general longitudinal autocorrelation structure which accommodates Gaussian-type AR(1), MA(1) and exchangeable correlations. After reviewing some existing methods of estimation (Liang and Zeger, 1986; Fitzmaurice and Laird, 1993; Hall and Severini, 1998; Sutradhar and Das, 1999), Sutradhar (2003) recommended GQL approach of estimation for the regression parameters as well as method of moments estimation of true longitudinal correlations in analyzing longitudinal discrete data. The method suggested by Sutradhar (2003) is much simpler than other existing methods and it does not encounter convergence problems.

Statistical techniques for analyzing longitudinal count data with excess zeros have been taking researcher's attention to a great extent in various disciplines like engineering, biomedical science, public health, demography, economics, and social science (Alfò and Maruotti, 2010; Mekonnen et al., 2019; Buu et al., 2012; Dobbie and Welsh, 2001; Min and Agresti, 2005; Zhu et al., 2017; Hasan and Sneddon, 2009; Hasan et al., 2016). In Chapter-3, a random effect MPois-Pois model with inference procedure has been developed for analyzing clustered data. In such cases, the intraclass correlation can be incorporated by considering a random effects term. A random effect model can also be applied in analyzing longitudinal data as it can incorporate the equicorrelation structure based lag correlations through the random effects (Sutradhar, 2011). Though random effect models are now frequently used in analyzing longitudinal data, the time effects cannot be accommodated by using these correlations. This is because the temporal effects of repeated measurement cannot be captured by using the fixed random effects for an individual (Sutradhar, 2011). To overcome this difficulty, we proposed a mixture of longitudinal Poisson count model in this chapter with a view to drawing marginal inference regarding exposure effects and longitudinal correlation for analyzing zero-inflated repeated measures count data.

Assuming various correlation structure such as AR(1), MA(1), equicorrelation structure, longitudinal data containing zero-heavy count responses have been studied by researchers in the past two decades (Hall and Zhang, 2004; Hasan and Sneddon, 2009; Hasan et al., 2016). In analyzing excess zero longitudinal count data, the researchers assumed existing method of zero-inflation i.e., they had considered that the data arise by the mixture of 'not-at-risk' and 'at-risk' populations. In order to analyze zero-inflated longitudinal count data, Hall and Zhang (2004) developed marginal models for Poisson part as well as for zero part separately. They employed the usual EM algorithm for fitting zero-inflated models, in which the M step is replaced by the solution of a GEE to take into account longitudinal correlation. Hasan and

Sneddon (2009) proposed an extension of ZIP model for analyzing longitudinal count data by utilizing a non stationary observation-driven time series model based correlation structure such as AR(1), MA(1), and equicorrelation structure. Following Sutradhar and Das (1999), they suggested to estimated the parameters of the proposed model using GQL technique by incorporating the true correlation structure. The main limitation of the observation-driven approach is that it cannot accommodate additional over-dispersion parameter which may present in longitudinal count data (Sutradhar and Bari, 2007; Hasan and Sneddon, 2009). This limitation is overcome by using a parameter-driven model, where the longitudinal correlation is captured through the latent process using random effects. A comparison of the observation-driven ZIP model (Hasan and Sneddon, 2009) with the parameter-driven model also have been studied by the researchers (Hasan et al., 2016).

A two-component mixture of 'at-risk' populations has been proposed for drawing marginal (marginalization over the subpopulations and over the individuals) inference in analyzing longitudinal count data in this study. For this purpose, we have extended the cross-sectional MPois-Pois mixture model (Benecha et al., 2017) given in Eq.(2.20)-Eq.(2.21) for drawing marginal inferences for longitudinal count data with excess zeros in this chapter. This model is named as repeated measures MPois-Pois (RMMPois-Pois) model. In the RMMPois-Pois model, it is assumed that if the observation obtained from an individual at the first time point belongs to a specific subpopulation, then all other observations obtained from that individual at other time points also belong to that subpopulation. A generalized quasilikelihood (GQL) based estimating equation approach for regression parameters and method of moment approach for true longitudinal correlation have been utilized following Sutradhar and Das (1999) for estimating the parameters of the proposed marginalized model.

4.1 MPois-Pois Model for Repeated Measures Count Data

Suppose that K independent individuals are selected randomly from a population and that the data are collected from the ith ($i = 1, \cdots, K$) individual at time points $t, t = 1, \cdots, T$. Let Y_{it} be the count variable of interest and $\boldsymbol{x}_{it}$ be a p-dimensional vector of covariates for the ith individual at tth time point. To develop the RMMPois-Pois model, it is assumed that the population is comprised of component-1 and component-2 and for the ith individual if the response y_{i1} is arisen from component-1 as the first repeated outcome, then for $t = 2, \cdots, T$, the set $\{y_{it}\}$ is also arisen from component-1; similarly, if y_{i1} comes from component-2, other responses $y_{i2}, \cdots, y_{iT}$ are also from component-2.

4.1.1 Two-component Mixture of Longitudinal Poisson Models

Let the mean response for component-1 at tth time point is $\mu_{1,it}$ and for component-2, it is denoted by $\mu_{2,it}, t = 1, \cdots, T$. Recall that the membership of the subpopulation is denoted by random variable D_i with the realization d_i, where D_i is the latent Bernoulli random variable with $P[D_i = 1] = \pi^*$. Then for $i = 1, \cdots, K$ and $t = 1, \cdots, T$, we have

$$Y_{it}|d_i \sim \text{Pois}\big(d_i\mu_{1,it} + [1 - d_i]\mu_{2,it}\big), d_i = 0, 1. \tag{4.1}$$

The conditional means, and variances are as follows

$$E\Big[Y_{it}|d_i\Big] = d_i\mu_{1,it} + (1 - d_i)\mu_{2,it} = \text{var}\Big[Y_{it}|d_i\Big].$$

Let ρ_l^* and ρ_l^{**} be the lag-$l, l = 1, \ldots, (T - 1)$ correlation between pair of observations obtained from any individual in component-1 and component-2, respectively. Then the conditional lag-l correlation is given as follows.

$$\text{corr}\Big[Y_{it}, Y_{it'}|d_i\Big] = d_i\rho_l^* + (1-d_i)\rho_l^{**}, l =\mid t-t' \mid, t \neq t', t, t' = 1, \cdots, T. \qquad (4.2)$$

Suppose that all possible lag-l correlations in the component-1 and the component-2 are represented by the correlation matrices $C_1(\rho^*)$ and $C_2(\rho^{**})$, respectively. Then $C_1(\rho^*)$ is defined as

$$C_1(\rho_1^*, \rho_2^*, \cdots, \rho_{T-1}^*) = \begin{bmatrix} 1 & \rho_1^* & \rho_2^* & \ldots & \rho_{T-1}^* \\ \rho_1^* & 1 & \rho_1^* & \ldots & \rho_{T-2}^* \\ \rho_2^* & \rho_1^* & 1 & \ldots & \rho_{T-3}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1}^* & \rho_{T-2}^* & \rho_{T-3}^* & \ldots & 1 \end{bmatrix}$$

and $C_2(\rho^{**})$ can be defined similarly.

Suppose that $\boldsymbol{Y}_i = (Y_{i1}, \cdots, Y_{it}, \cdots, Y_{iT})'$ be the $T \times 1$ random vector for the response variable and $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{it}, \cdots, \boldsymbol{x}_{iT})'$ be the $T \times p$ matrix of covariates for ith individual, $i = 1, \cdots, K$. Conditioning on the membership of the ith population, let $\boldsymbol{\mu}_{1,i} = (\mu_{1,i1}, \cdots, \mu_{1,it}, \cdots, \mu_{1,iT})'$, where $E[Y_{it}|d_i = 1] = \mu_{1,it}, t = 1, \cdots, T$ and $\boldsymbol{\mu}_{2,i} = (\mu_{2,i1}, \cdots, \mu_{2,it}, \cdots, \mu_{2,iT})'$, where $E[Y_{it}|d_i = 0] = \mu_{2,it}, t = 1, \cdots, T$. Note that both $\boldsymbol{\mu}_{1,i}$ and $\boldsymbol{\mu}_{2,i}, i = 1, \cdots, K$ are the $T \times 1$ vectors. Then the longitudinal Poisson-Poisson mixture model can be formed as

$$\log(\boldsymbol{\mu}_{1,i}) = \boldsymbol{X}_i\boldsymbol{\alpha}, \quad \text{and} \quad \log(\boldsymbol{\mu}_{2,i}) = \boldsymbol{X}_i\boldsymbol{\gamma}, i = 1, \cdots, K, \qquad (4.3)$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ both are the $p \times 1$ vector of parameters for component-1 mean and component-2 mean, respectively.

4.1.2 Marginalized Mixture of Two-Component Longitudinal Poisson Models

Suppose that the marginalized (unconditional) mean vector is denoted by $\boldsymbol{\mu}_i = E(\boldsymbol{Y}_i) = (\mu_{i1}, \cdots, \mu_{it}, \cdots, \mu_{iT})'$, where $E[Y_{it}] = \mu_{it}, t = 1, \cdots, T$. Using similar computations as in Eq.(2.15), the marginalized mean can be computed as

$$\boldsymbol{\mu}_i = [\pi^* \boldsymbol{\mu}_{1,i} + (1 - \pi^*)\boldsymbol{\mu}_{2,i}], i = 1, \cdots, K. \tag{4.4}$$

Then the elements of the component-2 mean vector $\boldsymbol{\mu}_{2,i}$ can be computed as

$$\mu_{2,it} = \frac{\mu_{it} - \pi^* \mu_{1,it}}{1 - \pi^*}, t = 1, \cdots, T. \tag{4.5}$$

Also, the marginalized (unconditional) variance at a specific time point of ith individual can be computed by using similar calculation as in Eq.(2.19), which is given by

$$\text{var}(Y_{it}) = \mu_{it} + \left[\frac{\pi^*}{1 - \pi^*}\right](\mu_{it} - \mu_{1,it})^2, i = 1, \cdots, K, t = 1, \cdots, T. \tag{4.6}$$

Suppose that all possible lag-l correlations in the marginalized (unconditional) population is represented by the correlation matrix $C(\rho)$, where

$$C(\rho) = C(\rho_1, \rho_2, \cdots, \rho_{T-1}) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1. \end{bmatrix}$$

The elements of $C(\rho)$ matrix, marginalized (unconditional) lag-l correlation between pair of observations obtained from any individual, can be defined as

$$\rho_l = \frac{\text{cov}(Y_{it}, Y_{it'})}{\sqrt{\text{var}(Y_{it})\text{var}(Y_{it'})}}, \forall i = 1, \cdots, K, l =\mid t - t' \mid, t \neq t', t, t' = 1, \cdots, T. \tag{4.7}$$

Following Benecha et al. (2017), we can model both the marginal mean vector $\boldsymbol{\mu}_i$ and component-1 mean vector $\boldsymbol{\mu}_{1,i}$ of Eq.(4.4) as

$$\log(\boldsymbol{\mu}_i) = \boldsymbol{X}_i \boldsymbol{\beta}, \quad \text{and} \quad \log(\boldsymbol{\mu}_{1,i}) = \boldsymbol{X}_i \boldsymbol{\alpha}, \tag{4.8}$$

respectively; where $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ both are the $p \times 1$ vector of parameters for marginalized mean and component-1 mean, respectively.

4.2 Estimation: GLM in Longitudinal Setup

The proposed model can be considered a mixture of GLMs in longitudinal setup, where the main interest is to draw marginal (marginalization over the subpopulations) inference of regression parameters. The most popular techniques for estimating the parameters of GLM in analyzing the longitudinal data are the GEE and GQL approaches. Both these techniques were derived from the quasi-likelihood (QL) estimation approach(Wedderburn, 1974; McCullagh, 1983). The following two subsections provide a brief discussion about GEE and GQL methods.

4.2.1 Generalized Estimating Equation Method

Let $Y_{it}, t = 1, \cdots, T, i = 1, \cdots, K$ be the response variable of interest and $\boldsymbol{x}_{it}$ be a p-dimensional vector of covariates for the ith individual at the tth time point. Assume that the probability distribution of Y_{it} is a member of an exponential family given as follows

$$f(y_{it}) = \exp\left[\{y_{it}\theta_{it} - b(\theta_{it}) + a(y_{it})\}\phi\right], i = 1, \cdots, K, t = 1, \cdots, T, \tag{4.9}$$

where $\theta_{it} = h(\eta_{it})$ with $\eta_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta}$, $\boldsymbol{\beta}$ be a $p \times 1$ vector of regression parameters, and ϕ is a possibly unknown scale parameter. If the underlying data strictly follow some probability model, ϕ may be assumed to known. For example, we may use $\phi = 1$ for binary and Poisson data. The first two moments of the probability distribution as shown in Eq.(4.9) are given by

$$E(Y_{it}) = \mu_{it} = b'(\theta_{it}), \text{var}(Y_{it}) = \sigma^2_{it} = \phi^{-1}b''(\theta_{it}) \tag{4.10}$$

where $b'(\theta_{it})$, and $b''(\theta_{it})$ are the first and second derivatives of $b(\theta_{it})$ with respect to θ_{it}, respectively. In longitudinal data setup, the repeated measurements of ith individual, $\boldsymbol{y}_i$, are correlated. In such instance, following Wedderburn (1974) the quasi-likelihood estimating equations can be written as

$$\sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \Sigma_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{4.11}$$

where $\boldsymbol{\mu}_i = E(\boldsymbol{Y}_i) = (\mu_{i1}, \cdots, \mu_{it}, \cdots, \mu_{iT})'$ is the mean vector and $\Sigma_i = \text{Var}(\boldsymbol{Y}_i)$ is the $T \times T$ covariance matrix of $\boldsymbol{y}_i$. But for discrete responses, the joint probability distribution of repeated observations have complicated functional form. Therefore, it is almost impossible to compute the covariance terms Σ_i in Eq.(4.11). To solve this difficulty, Liang and Zeger (1986) proposed a 'working' correlation structure based generalized estimating equation (GEE) for repeated measures data to obtain the consistent estimates of regression parameters for GLM.

Let $R(\zeta)$ be a $T \times T$ symmetric matrix fulfilling the requirement of being a correlation matrix, and let ζ a correlation parameter. Liang and Zeger (1986) refer the matrix, $R(\zeta)$ as a 'working' correlation matrix and suggested to use $\boldsymbol{V}_i = \boldsymbol{A}_i^{1/2} R(\zeta) \boldsymbol{A}_i^{1/2}$ instead of Σ_i in Eq.(4.11), where $\boldsymbol{A}_i = \text{diag}[\text{var}(Y_{i1}), \cdots, \text{var}(Y_{iT})]$. The $\boldsymbol{V}_i$ will be equal to Σ_i if $R(\zeta)$ is indeed the true correlation matrix of $\boldsymbol{y}_i$. Then the 'working' correlation based estimating equations suggested by Liang and Zeger (1986) are as follows

$$\sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \left[\boldsymbol{A}_i^{1/2} R(\zeta) \boldsymbol{A}_i^{1/2}\right]^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}. \tag{4.12}$$

To solve the estimating equations as given in Eq.(4.12), ζ is replaced by its consistent estimator (method of moment estimator), $\hat{\zeta}$. The method of moment estimator of ζ is computed by using Pearson residuals. The moment estimator of ζ is available in Liang and

Zeger (1986).

Under mild regularity conditions, the random quantity $K^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically multivariate normal with zero mean vector and covariance matrix V_T, where V_T is given as

$$V_T = \lim_{K \to \infty} K \left\{ \sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'} \right\}^{-1}. \tag{4.13}$$

4.2.2 Generalized Quasi-Likelihood Method

The GEE method of estimation for various choice of 'working' correlation matrix generally provides less efficient estimates than under the assumption of independent observations (Sutradhar and Das, 1999). With a view to increasing efficiency, Sutradhar and Das (1999) suggested a generalized quasi-likelihood (GQL) method of estimation for the regression parameters under GLM in longitudinal setup and the method of moments estimation for the associated true longitudinal correlations. The proposed GQL estimators of the regression parameter provide gain in efficiency as it uses true correlation structure.

To estimate the regression parameters for discrete data, Sutradhar and Das (1999) proposed to replace $R(\zeta)$ with the true correlation matrix $C(\rho)$ in Eq.(4.12), where $C(\rho)$ is the $T \times T$ symmetric matrix of true correlation defined as

$$C(\rho_1, \rho_2, \cdots, \rho_{T-1}) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{T-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \dots & 1 \end{bmatrix}$$

Then the quasi-likelihood estimator of $\boldsymbol{\beta}$ can be obtained as the root of the following estimating equations

$$\sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \left[\boldsymbol{A}_i^{1/2} C(\rho) \boldsymbol{A}_i^{1/2}\right]^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{4.14}$$

where $\boldsymbol{A}_i = \text{diag}[\text{var}(Y_{i1}), \cdots, \text{var}(Y_{iT})]$. As the solution of Eq.(4.14) requires the unknown true correlation matrix $C(\rho)$, Sutradhar and Kovacevic (2000) proposed the method of moment estimators of the elements of $C(\rho)$ as

$$\hat{\rho}_l = \frac{\sum_{i=1}^{K} \sum_{t=1}^{T-l} \tilde{y}_{it} \tilde{y}_{i,t+l} / K(T-l)}{\sum_{i=1}^{K} \sum_{t=1}^{T} \tilde{y}_{it}^2 / KT}, l = 1, \cdots, T-1, \tag{4.15}$$

where $\tilde{y}_{it}$ is the standardized residual defined as $\tilde{y}_{it} = (y_{it} - \mu_{it}) \big/ \sqrt{\text{var}(Y_{it})}$.

Suppose we denote the solution of the Eq.(4.14) as $\hat{\boldsymbol{\beta}}_{GQL}$. It can be shown that under mild regularity conditions, the random quantity $K^{1/2}(\hat{\boldsymbol{\beta}}_{GQL} - \boldsymbol{\beta})$ is asymptotically multivariate normal with zero mean vector and covariance matrix V_G. One can computes V_G as

$$V_G = \lim_{K \to \infty} K \left\{ \sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \left[\boldsymbol{A}_i^{1/2} C(\rho) \boldsymbol{A}_i^{1/2}\right]^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'} \right\}^{-1}. \tag{4.16}$$

The estimator $\hat{\boldsymbol{\beta}}_{GQL}$ is then consistent for $\boldsymbol{\beta}$ as well as more efficient than the estimator obtained by independent estimating equations.

4.3 GQL Estimation: RMMPois-Pois Model

Suppose that the population is comprised of component-1 and component-2 and if the first observation of ith individual i.e., y_{i1} belongs to component-$m (m = 1, 2)$, then for $t = 2, \cdots, T$, the set $\{y_{it}\}$ also belongs to component-m. In analyzing longitudinal data, the main objective is to propose a GQL estimate of the regression parameter of marginal (marginalization over the subpopulations) means, $\boldsymbol{\beta}$ as given in Eq.(4.8) under the repeated measures Poisson-Poisson mixture model. From Eq.(4.14) and Eq.(4.15), we have observed that the GQL approach requires computation of the quantity, marginal mean $E(Y_{it})$ and marginal variance $\text{Var}(Y_{it})$. Also from Eq.(4.6), we found that the quantity $\text{Var}(Y_{it})$ involves the mixing proportion (π^*) and the regression parameters, parameters of marginal means as well as component-1 means. Therefore, before implementing the GQL approach for estimating marginal parameters, it is required to have estimates of group membership of ith individual and also the nuisance parameter π^* (the mixing proportion).

In the mixture model setup of the repeated measurement, the data can be treated as incomplete because it involves unobserved variable for identifying group membership. However, it is possible to compute the conditional expectation of unobserved binary variable given the data in EM framework (discussed in Subsection-4.3.1) assuming that observations obtained from same individual are independent. As it is essential to identify the group membership of the mixture components to estimate the regression parameters of component-1, we proposed to use EM algorithm (Dempster et al., 1977) for estimating the expected value of latent binary variable for identifying group membership of each individual. Using these expected values an explicit formula for estimating mixing proportion (π^*) can be derived (discussed in Subsection-4.3.1). After getting the estimates, the regression parameters of the marginal model as well as the regression parameters of the component-1 can be estimated by using the GQL method. The true longitudinal correlation parameters for the marginal model and the component-1 model can be estimated by the method of moments. The estimation of parameters of interest can be summarized in following steps.

1. Estimate the group membership of ith individual, the expected value of the latent

variable D_i and the nuisance parameter π^* using EM algorithm.

2. Consider the initial value for the regression parameters of the marginal model, $\boldsymbol{\beta}$ and the regression parameters of the component-1, $\boldsymbol{\alpha}$.

3. Estimate the true correlation parameters for marginal model and the true correlation parameters for component-1 using method of moment approach.

4. Update the regression parameters of the marginal model, $\boldsymbol{\alpha}$ and the regression parameters of the component-1, $\boldsymbol{\beta}$ by using GQL approach.

5. Repeat Step-3 to Step-4 until convergence.

4.3.1 Estimating Mixing Proportion

We have assumed that if the response y_{i1} is arisen from component-m as the first repeated outcome, then for $t = 2, \cdots, T$, the set $\{y_{it}\}$ is also arisen from component-$m, m = 1, 2$ for the ith individual, $i = 1, \cdots, K$. Therefore, each $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{iT})', i = 1, \cdots, K$ can be thought of having arisen from only one of the components of the mixture model as represented in Eq.(4.1) for which we are intended to find estimates of associated marginal parameters as given in Eq.(4.8). In the EM framework of Poisson-Poisson mixture repeated data, the observed data vector within the formulation of the mixture problem is viewed as being incomplete i.e., the membership in the population, $d_{i.} = d_{it}$ for all $t = 1, \cdots, T$ is considered missing. As the individuals are independent and the membership of them in the population is constant with respect to occasions, it is assumed that the observations obtained from an individual over all the occasions are independent for the purpose of estimating the mixing probability, π^*. Further we denote $\boldsymbol{y} = (y_{11}, \cdots, y_{1T}, \cdots, y_{K1}, \cdots, y_{KT})'$ and $\boldsymbol{d} = (d_{11}, \cdots, d_{1T}, \cdots, d_{K1}, \cdots, d_{KT})'$. In order to estimating the nuisance parameters, the mixing proportion and the group membership, we then formulate the likelihood function for complete data under the assumption of independent observations as

$$L_c(\pi^*, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\boldsymbol{d}, \boldsymbol{y}) = \prod_{i=1}^{K}\prod_{t=1}^{T}\left[\pi^* \frac{e^{-\mu_{1,it}}\mu_{1,it}^{y_{it}}}{y_{it}!}\right]^{d_{it}}\left[(1-\pi^*)\frac{e^{-\mu_{2,it}}\mu_{2,it}^{y_{it}}}{y_{it}!}\right]^{1-d_{it}}. \tag{4.17}$$

The corresponding log likelihood is given by

$$\begin{aligned}\ell_c(\pi^*, \boldsymbol{\alpha}, \boldsymbol{\gamma}) =& \sum_{i=1}^{K}\sum_{t=1}^{T}\Big[d_{it}\text{logit}(\pi^*) + \log(1-\pi^*)\Big] \\ &+ \sum_{i=1}^{K}\sum_{t=1}^{T} d_{it}\Big[y_{it}\boldsymbol{x'_{it}\alpha} - \exp(\boldsymbol{x'_{it}\alpha}) - \log(y_{it}!)\Big] \\ &+ \sum_{i=1}^{K}\sum_{t=1}^{T}(1-d_{it})\Big[y_{it}\boldsymbol{x'_{it}\gamma} - \exp(\boldsymbol{x'_{it}}\gamma) - \log(y_{it}!)\Big].\end{aligned} \tag{4.18}$$

The EM algorithm is applied to this problem by treating the d_{it} as missing, which proceeds iteratively in two steps, E (for expectation) and M (for maximization).

E-Step

In EM framework, the latent data is handled by the E-step. It computes the conditional expectation of the complete data log likelihood, $\ell_c(\pi, \boldsymbol{\alpha}, \boldsymbol{\gamma})$, given the observed data $\boldsymbol{y}$, using the current fit $\boldsymbol{\theta}^{(r)}$ at the rth iteration, where $\boldsymbol{\theta} = (\pi^*, \boldsymbol{\alpha}', \boldsymbol{\gamma}')'$. Therefore, given the initial values $\boldsymbol{\theta}^{(0)} = (\pi^{*(0)}, \boldsymbol{\alpha'}^{(0)}, \boldsymbol{\gamma'}^{(0)})'$, the E-step computes

$$\begin{aligned}E\Big[\ell_c(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)})\Big] =& \sum_{i=1}^{K}\sum_{t=1}^{T}\Big(E\big[D_{it}|\boldsymbol{\theta}^{(0)}, y_{it}, \boldsymbol{x}_{it}\big]\text{logit}(\pi^*) + \log(1-\pi^*)\Big) \\ &+ \sum_{i=1}^{K}\sum_{t=1}^{T} E\Big[D_{it}|\boldsymbol{\theta}^{(0)}, y_{it}, \boldsymbol{x}_{it}\Big]\Big[y_{it}\boldsymbol{x}'_{it}\boldsymbol{\alpha} - \exp(\boldsymbol{x}'_{it}\boldsymbol{\alpha}) - \log(y_{it}!)\Big] \\ &+ \sum_{i=1}^{K}\sum_{t=1}^{T}\Big[1 - E[D_{it}|\boldsymbol{\theta}^{(0)}, y_{it}, \boldsymbol{x}_{it}]\Big]\Big[y_{it}\boldsymbol{x}'_{it}\boldsymbol{\gamma} - \exp(\boldsymbol{x}'_{it}\boldsymbol{\gamma}) - \log(y_{it}!)\Big].\end{aligned} \tag{4.19}$$

It follows that on the $(r+1)$th iteration, E-step requires the calculation of $E\Big[\ell_c(\pi^*, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\boldsymbol{\theta}^{(r)})\Big]$, where $\boldsymbol{\theta}^{(r)}$ is the value of $\boldsymbol{\theta}$ after the rth EM iteration. As the complete data log likelihood in Eq.(4.18) is linear in the unobservable data d_{it}, the E-step simply requires the calculation of the current conditional expectation of D_{it}, given the observed data $\boldsymbol{y}$. Now

$$\begin{aligned} E_{\boldsymbol{\theta}^{(r)}}(D_{it}|\boldsymbol{y}) &= Pr\Big[D_{it} = 1|\boldsymbol{y}\Big] \\ &= \frac{\pi^{*(0)}\frac{e^{-\mu_{1,it}}\mu_{1,it}^{y_{it}}}{y_{it}!}}{\pi^{*(0)}\frac{e^{-\mu_{1,it}}\mu_{1,it}^{y_{it}}}{y_{it}!} + (1-\pi^{*(0)})\frac{e^{-\mu_{2,it}}\mu_{2,it}^{y_{it}}}{y_{it}!}} \\ &\equiv P_{it}^{(0)}. \end{aligned} \tag{4.20}$$

Therefore, computation of the conditional expectation in Eq.(4.19) yields

$$\begin{aligned} E\Big[\ell_c(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)})\Big] &= \Big[\sum_{i=1}^{K}\sum_{t=1}^{T} P_{it}^{(0)}\text{logit}(\pi^*) + n\log(1-\pi^*)\Big] \\ &+ \Big[\sum_{i=1}^{K}\sum_{t=1}^{T} P_{it}^{(0)}\Big(y_{it}\boldsymbol{x}_{it}'\boldsymbol{\alpha} - \exp(\boldsymbol{x}_{it}'\boldsymbol{\alpha}) - \log(y_{it}!)\Big)\Big] \\ &+ \Big[\sum_{i=1}^{K}\sum_{t=1}^{T} \Big(1-P_{it}^{(0)}\Big)\Big(y_{it}\boldsymbol{x}_{it}'\boldsymbol{\gamma} - \exp(\boldsymbol{x}_{it}'\boldsymbol{\gamma}) - \log(y_{it}!)\Big)\Big] \\ &= \ell_{\pi^*} + \ell_{\boldsymbol{\alpha}} + \ell_{\boldsymbol{\gamma}}. \end{aligned} \tag{4.21}$$

M-Step

The M-step on the $(r+1)$th iteration requires the global maximization of Eq.(4.21) with respect to $\boldsymbol{\theta}$ over the parameter space $\boldsymbol{\Omega}$ to give the update estimate $\boldsymbol{\theta}^{(r+1)}$. For the Poisson mixture, the update $\boldsymbol{\theta}^{(r+1)} = (\pi^{*(r+1)}, \boldsymbol{\alpha}'^{(r+1)}, \boldsymbol{\gamma}'^{(r+1)})'$ can be obtained separately as three components $\ell_{\pi}^*, \ell_{\boldsymbol{\alpha}}$ and $\ell_{\boldsymbol{\gamma}}$ in Eq.(4.21) can be optimized separately. Maximizing ℓ_{π}^* with respect to π^* of Eq.(4.21) gives the updated estimate of π^* as

$$\pi^{*(1)} = \frac{\sum_{i=1}^{K}\sum_{t=1}^{T} P_{it}^{(0)}}{KT}. \tag{4.22}$$

Components $\ell_{\boldsymbol{\alpha}}$ and $\ell_{\boldsymbol{\gamma}}$ of Eq.(4.21) correspond to weighted log likelihood of GLM so that they can be solved iteratively using Fisher's method of scoring to obtain the next set of parameters $\boldsymbol{\alpha}^{(1)}$ and $\boldsymbol{\gamma}^{(1)}$. For a GLM, it is equivalent to using iteratively reweighed least squares (IRLS) (McLachlan and Peel, 2000).

As EM-algorithm always provides convergence sequence of estimator, the E- and M-steps are continued repeatedly until convergence. If the convergence is achieved at the hth step, the estimates of conditional expectation of group membership, $P_{it}^{(h)}$ and the mixing proportion, $\pi^{*(h)}$ are required for implementing the GQL approach. The EM-estimates of the regression parameters $\boldsymbol{\alpha}^{(h)}$ and $\boldsymbol{\gamma}^{(h)}$ are not of our interest as these are estimated using the assumption of independent measurements. Hence, these are not used further in the estimation approach.

4.3.2 Estimation of Regression Parameters: Marginal Model

Let $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{it}, \cdots, y_{iT})'$ be a $T \times 1$ vector of repeated count observations obtained from ith individual. Again let $\boldsymbol{\mu}_i = (\mu_{i1}, \cdots, \mu_{iT})'$ be the $T \times 1$ mean vector and $\Sigma_i = \text{Var}(\boldsymbol{Y}_i)$ be the $T \times T$ covariance matrix. But for repeated discrete responses, it is almost impossible to compute Σ_i because of its complicated multivariate distribution. One may express the Σ_i as $\Sigma_i(\rho) = \boldsymbol{A}_i^{1/2} C(\rho) \boldsymbol{A}_i^{1/2}$, where $\boldsymbol{A}_i = \text{diag}[\text{var}(Y_{i1}), \cdots, \text{var}(Y_{iT})]$ be the $T \times T$ diagonal matrix of unconditional variance (the elements of $\boldsymbol{A}_i$ are shown in Eq.(4.6)) and $C(\rho) = C(\rho_1, \cdots, \rho_{T-1})$ be the $T \times T$ true correlation matrix $i = 1, 2, \cdots, K$, where the true lag-l correlation of the observations is represented by $\rho_l, l = |t - t'|, t \neq t', t, t' = 1, \cdots, T$. For a known correlation matrix $C(\rho)$, the GQL estimating equation (Sutradhar,

2003) for the regression parameters $\boldsymbol{\beta}$ in the marginalized model can be written as

$$\sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \Sigma_i^{-1}(\rho)(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{4.23}$$

where $\frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} = \boldsymbol{X}_i' \mathbb{Z}_i$ with $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{it}, \cdots, \boldsymbol{x}_{iT})'$ is a $T \times p$ matrix and $\mathbb{Z}_i = \text{diag}[\mu_{i1}, \cdots, \mu_{iT}]$ is a $T \times T$ matrix. The corresponding GQL estimates of $\boldsymbol{\beta}$ can be computed iteratively by using Newton-Raphson method. Thus the estimate of $\boldsymbol{\beta}$ in the $(r+1)$th iteration can be expressed as

$$\hat{\boldsymbol{\beta}}_{GQL}^{(r+1)} = \hat{\boldsymbol{\beta}}_{GQL}^{(r)} + \Big[\sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \Sigma_i^{-1}(\hat{\rho}) \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'}\Big]^{-1} \Big[\sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \Sigma_i^{-1}(\hat{\rho})(\boldsymbol{y}_i - \boldsymbol{\mu}_i)\Big]. \tag{4.24}$$

Under some mild regularity conditions, it can be shown that the quantity $K^{1/2}(\hat{\boldsymbol{\beta}}_{GQL} - \boldsymbol{\beta})$ is asymptotically multivariate normal with zero mean vector and covariance matrix V^* (Sutradhar, 2003), where V^* is given as

$$V^* = \lim_{K \to \infty} K \left\{\sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \Sigma_i^{-1}(\hat{\rho}) \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'}\right\}^{-1}. \tag{4.25}$$

4.3.3 Estimation of Correlation Parameter: Marginal Model

If the true correlation matrix $C(\rho)$ is unknown, one may estimates the elements of the correlation matrix by the method of moment (Sutradhar and Kovacevic, 2000). For $l = 1, \cdots, T-1$, the method of moments estimator of the true correlation of lag l, ρ_l for the marginalized model is given as

$$\hat{\rho}_l = \frac{\sum_{i=1}^{K} \sum_{t=1}^{T-l} \tilde{y}_{it} \tilde{y}_{i,t+l} / K(T-l)}{\sum_{i=1}^{K} \sum_{t=1}^{T} \tilde{y}_{it}^2 / KT}, \tag{4.26}$$

where $\tilde{y}_{it}$ is the standardized residual defined as

$$\tilde{y}_{it} = (y_{it} - \mu_{it}) / \sqrt{\text{var}(Y_{it})}, \tag{4.27}$$

where μ_{it} is given in Eq.(4.8) and $\text{var}(Y_{it})$ is given in Eq.(4.6).

4.3.4 Estimating Regression Parameters: Component-1 Model

The E-step of EM algorithm as given in Eq.(4.20) provides the expected value (conditional on the given data) of unobserved binary variable $D_{it}, t = 1, \cdots, T$ for identifying group membership of ith individual. If convergency is achieved at the hth step, the estimates of conditional expectation of the unobserved binary variable is denoted by $P_{it}^{(h)}$. Then the expected value of the unobserved binary variable for the ith individual can be computed as $D_{i.} = \frac{\sum_{t=1}^{T} P_{it}^{(h)}}{T}, i = 1, \cdots, K, \forall t = 1, \cdots, T$. An individual can be identified as member of component-1 if $D_{i.} \geq 0.50$; otherwise, the individual is identified as member of component-2.

Suppose that we have identified K_1 individuals as member of component-1 and K_2 individuals as member of component-2 by using the above classification approach. The individuals from both groups comprise the whole sample i.e, $K = K_1 + K_2$. Let, an individual $i, i = 1, \cdots, K_1$ is observed at time points $t, t = 1, \cdots, T$ and at each t we have a scalar response $Y_{1,it}$ and a p-dimensional vector of covariates $\boldsymbol{x}_{1,it}$ in the component-1 of mixture. For the ith individual, let $\boldsymbol{y}_{1,i} = (y_{1,i1}, \cdots, y_{1,iT})'$ be the response vector and $\boldsymbol{X}_{1,i} = (\boldsymbol{x}_{1,i1}, \cdots, \boldsymbol{x}_{1,it}, \cdots, \boldsymbol{x}_{1,iT})'$ be the $T \times p$ matrix of covariates. Also let $\text{var}(Y_{1,it}) = E(Y_{1,it}) = \mu_{1,it}$, where $\mu_{1,it}$ is given in Eq.(4.8). The variance-covariance matrix of $Y_{1,i1}$ is given by

$$\Sigma_{1,i}(\rho^*) = \boldsymbol{A}_{1,i}^{1/2} C_1(\rho^*) \boldsymbol{A}_{1,i}^{1/2},$$

where $A_{1,i} = \text{diag}[\text{var}(Y_{1,i1}), \cdots, \text{var}(Y_{1,iT})]$ is the $T \times T$ diagonal matrix of variance component of longitudinal Poisson model and $C_1(\rho^*)$ is the true longitudinal correlation for component-1. For known true correlation matrix $C_1(\rho^*)$, the GQL estimating equation (Sutradhar, 2003) for the regression parameters for model of component-1 then can be written as

$$\sum_{i=1}^{K_1} \boldsymbol{X}_{1,i}^{\top} \boldsymbol{A}_{1,i} \Sigma_{1,i}^{-1}(\hat{\rho}^*)(\boldsymbol{Y}_{1,i} - \boldsymbol{\mu}_{1,i}) = \mathbf{0}$$

$$\Rightarrow \sum_{i=1}^{K_1} \boldsymbol{X}_{1,i}^{\top} \boldsymbol{A}_{1,i}^{1/2} C_1^{-1}(\rho^*) \boldsymbol{A}_{1,i}^{-1/2}(\boldsymbol{y}_{1,i} - \boldsymbol{\mu}_{1,i}) = \mathbf{0}. \tag{4.28}$$

The estimate of component-1 regression parameters $\boldsymbol{\alpha}$ at the $(r+1)$th iteration obtained by using the Newton-Raphson algorithm can be expressed as

$$\hat{\boldsymbol{\alpha}}_{GQL}^{(r+1)} = \hat{\boldsymbol{\alpha}}_{GQL}^{(r)} + \Big[\sum_{i=1}^{K_1} \boldsymbol{X}_{1,i}^{\top} \boldsymbol{A}_{1,i}^{1/2} C_1^{-1}(\rho^*) \boldsymbol{A}_{1,i}^{1/2} \boldsymbol{X}_{1,i}\Big]^{-1} \Big[\sum_{i=1}^{K_1} \boldsymbol{X}_{1,i}^{\top} \boldsymbol{A}_{1,i}^{1/2} C_1^{-1}(\rho^*) \boldsymbol{A}_{1,i}^{-1/2}(\boldsymbol{y}_{1,i} - \boldsymbol{\mu}_{1,i})\Big]. \tag{4.29}$$

Under some mild regularity conditions, it can be shown that the quantity $K_1^{-1/2}(\hat{\boldsymbol{\alpha}}_{GQL} - \boldsymbol{\alpha})$ is asymptotically multivariate normal with zero mean vector and covariance matrix V_1^*, where

$$V_1^* = \lim_{K_1 \to \infty} K_1 \Big\{\sum_{i=1}^{K_1} \boldsymbol{X}_{1,i}^{\top} \boldsymbol{A}_{1,i}^{1/2} C_1^{-1}(\rho^*) \boldsymbol{A}_{1,i}^{1/2} \boldsymbol{X}_{1,i}\Big\}^{-1}. \tag{4.30}$$

4.3.5 Estimation of Correlation Parameter: Component-1 Model

For unknown true correlation matrix $C_1(\rho^*)$ is , the elements of the correlation matrix can be estimated by the method of moment (Sutradhar and Kovacevic, 2000). For $l = 1, \cdots, T-1$, the method of moments estimator of the true correlation of lag l, $\hat{\rho}_l^*$ for the component-1 model is given as

$$\hat{\rho}_l^* = \frac{\sum_{i=1}^{K_1} \sum_{t=1}^{T-l} \tilde{y}_{1,it} \tilde{y}_{1i,t+l} / K_1(T-l)}{\sum_{i=1}^{K_1} \sum_{t=1}^{T} \tilde{y}_{1,it}^2 / K_1 T}, \tag{4.31}$$

where $\tilde{y}_{1,it}$ is the standardized residual defined as

$$\tilde{y}_{1,it} = (y_{1,it} - \mu_{1,it}) / \sqrt{\mu_{1,it}}, \tag{4.32}$$

where $\mu_{1,it}$ is given in Eq.(4.8).

4.4 Simulation Study

Extensive simulation studies were conducted to investigate the performance of the proposed RMMPois-Pois model. For this purpose, mixture of correlated Poisson data have been generated following three widely used AR(1), MA(1) and exchangeable autocorrelation structures. Simulated data were generated for different proportions of mixture (π^*), for different sample sizes (K) and for different numbers of occasions (T) using each of the three correlation structures. For instance, we had considered $K = 100, 200, 500$ each at varying number of occasions ($T = 3, 4, 5$). To obtain marginal inference from longitudinal Poisson-Poisson mixture distribution, the zero-inflated data had been generated using different $\pi^* = 0.50, 0.70, 0.90$. In the simulation studies, it was assumed that both the models (model for unconditional mean and component-1 mean) given in Eq.(4.8) were influenced by the same set of known covariates.

4.4.1 Marginalized Poisson AR(1)-Poisson AR(1) Mixture Probability Model

Let us consider the stationary AR(1) based Poisson model (McKenzie, 1988) for component-1 as

$$y_{it} = \rho_1 * y_{i,t-1} + w_{it}^{(1)}, i = 1, \cdots, K, t = 1, \cdots, T, \tag{4.33}$$

where $Y_{i,t-1} \sim \text{Pois}(\mu_{1,i.})$ with $\mu_{1,i.} = \exp(\boldsymbol{x}'_{i.}\boldsymbol{\alpha})$, $\boldsymbol{x}'_{i.} = \boldsymbol{x}'_{it}, \forall t = 1, \cdots, T$ (i.e., all covariates are time independent); ρ_1 is a constant scale parameter satisfying the range restriction $0 \leq \rho_1 \leq 1$. For a given $y_{i,t-1}$, $\rho_1 * y_{i,t-1}$ in Eq.(4.33) is computed through a binomial thinning operation (McKenzie, 1988). More specifically, $\rho_1 * y_{i,t-1}$ is the sum of $y_{i,t-1}$ binary random variables each with probability of success ρ_1. Mathematically we can write,

$$\rho_1 * y_{i,t-1} = \sum_{j=1}^{y_{i,t-1}} b_j(\rho_1) = z_{i,t-1}^{(1)}, \tag{4.34}$$

with $\Pr[b_j(\rho_1) = 1] = \rho_1$ and $\Pr[b_j(\rho_1) = 0] = 1 - \rho_1$. That is, $Z_{i,t-1}^{(1)}|y_{i,t-1} \sim \text{Bin}(y_{i,t-1}, \rho_1)$. Assume that $W_{it}^{(1)} \sim \text{Pois}(\mu_{1,i.}(1-\rho_1))$ with $W_{it}^{(1)}$ and $Z_{i,t-1}^{(1)}$ are independent. The mean of the ith individual at all the occasions can be computed as

$$\begin{aligned}
E\big[Y_{i1}\big] =& \mu_{1,i.}, \\
E\big[Y_{i2}\big] =& E\big[E\big(Y_{i2}|Y_{i1}\big)\big] = E\big[E\big(Z_{i1}^{(1)}|Y_{i1} + W_{i2}^{(1)}\big)\big] = E\big[Y_{i1}\rho_1 + (1-\rho_1)\mu_{1,i.}\big] = \mu_{1,i.}.
\end{aligned}$$

Similarly, we can compute $E\big[Y_{i3}\big] = \cdots = E\big[Y_{iT}\big] = \mu_{1,i.}$. The variance of the ith individual at all the occasions can be determined as

$$\begin{aligned}
\text{Var}\big[Y_{i1}\big] =& \mu_{1,i.}, \\
\text{Var}\big[Y_{i2}\big] =& \text{Var}\big[E\big(Y_{i2}|Y_{i1}\big)\big] + E\big[\text{Var}\big(Y_{i2}|Y_{i1}\big)\big] \\
=& \text{Var}\big[\rho_1 Y_{i1} + (1-\rho_1)\mu_{1,i.}\big] + E\big[\rho_1(1-\rho_1)Y_{i1} + (1-\rho_1)\mu_{1,i.}\big] = \mu_{1,i.}.
\end{aligned}$$

Similarly, one can compute $\text{Var}\big[Y_{i3}\big] = \cdots = \text{Var}\big[Y_{iT}\big] = \mu_{1,i.}$. The expected value of the lag-l product of Y_{it} for the ith individual can be calculated as

$$\begin{aligned}
E\big[Y_{it}Y_{i,t-1}\big] =& E\big[Y_{i,t-1}E\big(Y_{it}|Y_{i,t-1}\big)\big] = E\big[Y_{i,t-1}\big(Y_{i,t-1}\rho_1 + (1-\rho_1)\mu_{1,i.}\big)\big] = \mu_{1,i.}\rho_1 + \mu_{1,i.}^2, \\
E\big[Y_{it}Y_{i,t-2}\big] =& E\big[Y_{i,t-2}E\big\{E\big(Y_{it}|Y_{i,t-1}\big)|Y_{i,t-2}\big\}\big] \\
=& E\big[Y_{i,t-2}E\big\{Y_{i,t-1}\rho_1 + (1-\rho_1)\mu_{1,i.}|Y_{i,t-2}\big\}\big] \\
=& E\big[Y_{i,t-2}\big\{\rho_1 E\big(Y_{i,t-1}|Y_{i,t-2}\big) + (1-\rho_1)\mu_{1,i.}\big\}\big] \\
=& E\big[Y_{i,t-2}\big\{\rho_1\big(Y_{i,t-2}\rho_1 + (1-\rho_1)\mu_{1,i.}\big) + (1-\rho_1)\mu_{1,i.}\big\}\big] = \mu_{1,i.}\rho_1^2 + \mu_{1,i.}^2.
\end{aligned}$$

Similarly, we can calculate $E\big[Y_{it}Y_{i,t-l}\big] = \mu_{1,i.}\rho_1^l + \mu_{1,i.}^2$. The lag-$l$ covariance between Y_{it} and $Y_{i,t-l}$ can be found as

$$\text{Cov}(Y_{it}, Y_{i,t-l}) = \mu_{1,i.}\rho_1^l, l = 1, \cdots, T-1, t > l. \tag{4.35}$$

The lag-l correlation between Y_{it} and $Y_{i,t-l}$ for component-1 is then computed as

$$\rho_l^* = \text{corr}(Y_{it}, Y_{i,t-l}|d_i = 1) = \rho_1^l, l = 1, \cdots, T-1, t > l. \tag{4.36}$$

Again, let us consider consider the stationary AR(1) based Poisson model (McKenzie, 1988) for component-2 as

$$y_{it} = \rho_2 * y_{i,t-1} + w_{it}^{(2)}, i = 1, \cdots, K, t = 1, \cdots, T, \tag{4.37}$$

where $0 \leq \rho_2 \leq 1$, $Y_{i,t-1} \sim \text{Pois}(\mu_{2,i.})$ with $\mu_{2,i.} = \mu_{2,it} \forall t = 1, \cdots, T$. Note that $\mu_{2,i.}$ is defined in Eq.(4.5), and ρ_2 is defined similarly as ρ_1 for component-1. Like component-1, we have assumed that $\rho_2 * y_{i,t-1}$ in Eq.(4.37) can be computed through a binomial thinning operation as

$$\rho_2 * y_{i,t-1} = \sum_{j=1}^{y_{i,t-1}} b_j(\rho_2) = z_{i,t-1}^{(2)}.$$

Similarly as component-1, it can be shown that for component-2, $Y_{it} \sim \text{Pois}(\mu_{2,i.})$. Then the lag-$l$ correlation between Y_{it} and $Y_{i,t-l}$ for component-2 can be computed similarly as computed for component-1, which is $\rho_l^{**} = \rho_2^l$.

The Poisson AR(1)-Poisson AR(1) mixture probability model can be constructed then by using the conditional model as provided in Eq.(4.1), where Poisson AR(1) probability model for component-1 and component-2 are given in Eq.(4.33) and Eq.(4.37), respectively. In constructing the mixture model, we have assumed that when an individual provides an observation from component-$m, m = 1, 2$ at first occasion then all subsequent observations for that individual at subsequent occasions will also belong to the same component. Then the conditional mean, variance and lag-l covariances with respect to the group membership of the ith individual are as follows

$$E\Big[Y_{it}|d_i\Big] = d_i\mu_{1,i.} + (1-d_i)\mu_{2,i.}, \forall t = 1, \cdots, T,$$

$$\text{Var}\Big[Y_{it}|d_i\Big] = d_i\mu_{1,i.} + (1-d_i)\mu_{2,i.}, \forall t = 1, \cdots, T,$$

$$\text{Cov}\Big[Y_{it}, Y_{i,t-l}|d_i\Big] = d_i\rho_1^l\mu_{1,i.} + (1-d_i)\rho_2^l\mu_{2,i.}, l = 1, \cdots, T-1, \forall t > l.$$

The unconditional (marginalized over the subpopulations) mean, and variance are then computed by using Eq.(4.4), and Eq.(4.6), respectively. Also, the unconditional lag-l covariance under Poisson AR(1)-Poisson AR(1) mixture can be obtained as

$$\begin{aligned}
\text{Cov}\Big[Y_{it}, Y_{i,t-l}\Big] =& E\Big[\text{Cov}\big\{Y_{it}, Y_{i,t-l}|d_i\big\}\Big] + \text{Cov}\Big[E\big\{Y_{it}|d_i\big\}, E\big\{Y_{i,t-l}|d_i\big\}\Big] \\
=& E\Big[D_i\rho_1^l\mu_{1,i.} + (1-D_i)\rho_2^l\mu_{2,i.}\Big] + \\
& \text{Cov}\Big[D_i\mu_{1,i.} + (1-D_i)\mu_{2,i.}, D_i\mu_{1,i.} + (1-D_i)\mu_{2,i.}\Big] \\
=& \Big[\pi^*\rho_1^l\mu_{1,i.} + (1-\pi^*)\rho_2^l\mu_{2,i.}\Big] + \text{Var}\Big[D_i\mu_{1,i.} + (1-D_i)\mu_{2,i.}\Big] \\
=& \Big[\pi^*\rho_1^l\mu_{1,i.} + (1-\pi^*)\rho_2^l\mu_{2,i.}\Big] + \Big[\pi^*(1-\pi^*)\mu_{1,i.}^2 + \pi^*(1-\pi^*)\mu_{2,i.}^2 + \\
& 2\mu_{1,i.}\mu_{2,i.}\text{Cov}[D_i, 1-D_i]\Big] \\
=& \Big[\pi^*\rho_1^l\mu_{1,i.} + (1-\pi^*)\rho_2^l\mu_{2,i.}\Big] + \Big[\pi^*(1-\pi^*)\big\{\mu_{2,i.} - \mu_{1,i.}\big\}^2\Big] \\
=& \Big[\pi^*\rho_1^l\mu_{1,i.} + (1-\pi^*)\rho_2^l\mu_{2,i.}\Big] + \Big(\frac{\pi^*}{1-\pi^*}\Big)(\mu_{i.} - \mu_{1,i.})^2. \qquad (4.38)
\end{aligned}$$

We have assumed same probability of success for the binomial thinning operation in both the components of mixture, for simplicity , i.e., we use $\rho_1 = \rho_2 = \rho^*$. Then Eq.(4.38) becomes

$$\text{Cov}\Big[Y_{it}, Y_{i,t-l}\Big] = \rho^{*l}\mu_{i.} + \Big(\frac{\pi^*}{1-\pi^*}\Big)(\mu_{i.} - \mu_{1,i.})^2. \qquad (4.39)$$

Therefore, the unconditional lag-l correlation defined in Eq.(4.7) for the ith individual under marginalized Poisson AR(1)-Poisson AR(1) mixture model setup can be obtained from Eq.(4.6) and Eq.(4.39) as

$$\rho_{l_i} = \text{corr}\Big[Y_{it}, Y_{i,t-l}\Big] = \frac{\rho^{*l}\mu_{i.} + \left(\frac{\pi^*}{1-\pi^*}\right)(\mu_{i.} - \mu_{1,i.})^2}{\mu_{i.} + \left(\frac{\pi^*}{1-\pi^*}\right)(\mu_{i.} - \mu_{1,i.})^2}, l = 1, \cdots, T-1, i = 1, \cdots, K, \forall t > l. \tag{4.40}$$

Note that Eq.(4.40) reduces to correlation structure of Poisson AR(1) probability model (Sutradhar, 2003; McKenzie, 1988) if and only if $\mu_{1,i.} = \mu_{2,i.} = \mu_{i.}$. Finally, the longitudinal correlation under proposed marginalized model is obtained by averaging the K quantities computed from Eq.(4.40) as

$$\rho_l = \frac{1}{K}\sum_{i=1}^{K} \rho_{l_i}, l = 1, \cdots, T-1. \tag{4.41}$$

Data Generation from RMMPois-Pois Model: Mixture of Poisson AR(1) Processes

To obtain marginal inference from RMMPois-Pois model using mixture of Poisson AR(1) models, the zero-inflated data had been generated from K individuals each at T occasions by the following steps.

1. The covariate $\boldsymbol{x}_{it} = (x_{it1}, x_{it2})'$ were generated in such a way that x_{it1} ∼unif(0,1), and x_{it2} ∼Bernoulli(0.40), $i = 1, \cdots, K, t = 1$.

2. In Eq.(4.8), suitable values of the regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)' = (0.20, 0.60, 0.50)'$ were used to compute the marginal means at first occasion $\mu_{i1}, i = 1, \cdots, K$ and suitable values of the regression parameters $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)' = (-1.0, 0.40, 0.50)'$ were used to compute the component-1 means at first occasion $\mu_{1,i1}, i = 1, \cdots, K$.

3. The component-2 means at first occasion ($\mu_{2,i1}$) were then computed using Eq.(4.5).

4. Binary observations $d_i, i = 1, \cdots, K$ were generated using $D \sim$ Bernoulli(π^*).

5. Under RMMPois-Pois setup, zero-inflated counts from K individuals at first occasion, y_{i1} were then generated.

6. Other $T-1$ repeated observations from ith individual, $i =, 1, \cdots, K$ were generated using observation-driven Poisson model as given in either Eq.(4.33) or in Eq.(4.37) conditioning on the value of $d_i = 1$ or $d_i = 0$, respectively. The value of $\rho_1 = \rho_2 = \rho^* = 0.40$ and 0.70 had been considered while using the Eq.(4.33) and the Eq.(4.37).

7. For different values of K, T and for given values of $\boldsymbol{\beta}, \boldsymbol{\alpha}, \rho^*$ various proportion of zeros have been generated. For instances, data contain approximately 33% zeros for $\pi^* = 0.50$, approximately 43% zeros for $\pi^* = 0.70$, and approximately 54% zeros for $\pi^* = 0.90$.

4.4.2 Marginalized Poisson MA(1)-Poisson MA(1) Mixture Probability Model

Let us consider the stationary MA(1) based Poisson model (McKenzie, 1988) for component-1 as

$$y_{it} = \rho_1 * w^{(1)}_{i,t-1} + w^{(1)}_{it}, i = 1, \cdots, K, t = 1, \cdots, T, \tag{4.42}$$

where $W^{(1)}_{it} \overset{\text{iid}}{\sim} \text{Pois}\left(\frac{\mu_{1,i.}}{1+\rho_1}\right)$ with $\mu_{1,i.} = \exp(\boldsymbol{x}'_{i.}\boldsymbol{\alpha})$, $\boldsymbol{x}'_{i.} = \boldsymbol{x}'_{it}, \forall t = 1, \cdots, T$ (i.e., all covariates are time independent); ρ_1 is the parameter for longitudinal correlation satisfying the range restriction $0 \leq \rho_1 \leq 1$. Like AR(1) process, we have considered $\rho_1 * w^{(1)}_{i,t-1}$ in Eq.(4.42) is computed through a binomial thinning operation.

Following AR(1) model for component-1, one can derive the expression for mean and variance at all occasion, and expected value of the products for lag-l. The mean of the ith individual at all the occasions can be computed as

$$E\big[Y_{it}\big] = \mu_{1,i.} = \text{Var}\big[Y_{it}\big], i = 1, \cdots, K, t = 1, \cdots, T.$$

The expected value of the lag-l product of Y_{it} for the ith individual can be computed as

$$E\big[Y_{it}Y_{i,t-1}\big] = \mu_{1,i.}\frac{\rho_1}{1+\rho_1} + \mu_{1,i.}^2,$$

$$E\big[Y_{it}Y_{i,t-2}\big] = \mu_{1,i.}^2.$$

Similarly, we can calculate $E\big[Y_{it}Y_{i,t-l}\big] = \mu_{1,i.}^2, l > 2$. The lag-$l$ correlation for component-1, defined as in Eq.(4.2), can be obtained as

$$\rho_l^* = \text{corr}\Big[Y_{it}, Y_{i,t-l}|d_i = 1\Big] = \begin{cases} \dfrac{\rho_1}{1+\rho_1} & \text{for} \quad l = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.43}$$

Again, let us consider the stationary MA(1) based Poisson model (McKenzie, 1988) for component-2 as

$$y_{it} = \rho_2 * w_{i,t-1}^{(2)} + w_{it}^{(2)}, i = 1, \cdots, K, t = 1, \cdots, T, \tag{4.44}$$

where $W_{it}^{(2)} \overset{\text{iid}}{\sim} \text{Pois}\Big(\frac{\mu_{2,i.}}{1+\rho_2}\Big)$ with $\mu_{2,i.} = \mu_{2,it} \forall t = 1, \cdots, T$. Note that $\mu_{2,i.}$ is defined in Eq.(4.5), and ρ_2 is defined similarly as ρ_1 for component-1. By similar calculations as for component-1 MA(1) process, we can easily obtain $Y_{it} \sim \text{Pois}(\mu_{2,i.})$ for component-2 MA(1) process with lag-l correlation as

$$\rho_l^{**} = \text{corr}\Big[Y_{it}, Y_{i,t-l}|d_i = 0\Big] = \begin{cases} \dfrac{\rho_2}{1+\rho_2} & \text{for} \quad l = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The unconditional (marginalized over the subpopulations) mean, and variance are then computed by using Eq.(4.4), and Eq.(4.6), respectively. Similarly as Subsection-4.4.1, the unconditional lag-l covariance under Poisson MA(1)-Poisson MA(1) mixture can be obtained as

$$\text{Cov}\Big[Y_{it}, Y_{i,t-l}\Big] = \begin{cases} \Big[\pi^* \dfrac{\rho_1}{1+\rho_1}\mu_{1,i.} + (1-\pi^*)\dfrac{\rho_2}{1+\rho_2}\mu_{2,i.}\Big] + \Big(\dfrac{\pi^*}{1-\pi^*}\Big)(\mu_{i.} - \mu_{1,i.})^2 & \text{for} \quad l = 1 \\ \Big(\dfrac{\pi^*}{1-\pi^*}\Big)(\mu_{i.} - \mu_{1,i.})^2 & \text{otherwise.} \end{cases}$$

For simplicity, we have assumed $\rho_1 = \rho_2 = \rho^*$. Then the covariance between Y_{it} and $Y_{i,t-l}$ becomes

$$\text{Cov}\Big[Y_{it}, Y_{i,t-l}\Big] = \begin{cases} \Big(\dfrac{\rho^*}{1+\rho^*}\Big)\mu_{i.} + \Big(\dfrac{\pi^*}{1-\pi^*}\Big)(\mu_{i.} - \mu_{1,i.})^2 & \text{for} \quad l = 1 \\ \Big(\dfrac{\pi^*}{1-\pi^*}\Big)(\mu_{i.} - \mu_{1,i.})^2 & \text{otherwise.} \end{cases}$$

Therefore, the unconditional lag-l correlation defined in Eq.(4.7) for the ith individual under marginalized Poisson MA(1)-Poisson MA(1) mixture model setup can be obtained as

$$\rho_{l,i} = \text{corr}\Big[Y_{it}, Y_{i,t-l}\Big] = \begin{cases} \dfrac{\Big(\frac{\rho^*}{1+\rho^*}\Big)\mu_{i.} + \Big(\frac{\pi^*}{1-\pi^*}\Big)(\mu_{i.} - \mu_{1,i.})^2}{\mu_{i.} + \big(\frac{\pi^*}{1-\pi^*}\big)(\mu_{i.} - \mu_{1,i.})^2} & \text{for} \quad l = 1 \\ \dfrac{\Big(\frac{\pi^*}{1-\pi^*}\Big)(\mu_{i.} - \mu_{1,i.})^2}{\mu_{i.} + \big(\frac{\pi^*}{1-\pi^*}\big)(\mu_{i.} - \mu_{1,i.})^2} & \text{otherwise.} \end{cases} \tag{4.45}$$

Note that Eq.(4.45) reduces to correlation structure of Poisson MA(1) probability model (Sutradhar, 2003; McKenzie, 1988) if and only if $\mu_{1,i.} = \mu_{2,i.} = \mu_{i.}$. Finally, the longitudinal correlation under proposed marginalized model is obtained by averaging the K quantities of Eq.(4.45) using Eq.(4.41).

Data Generation from RMMPois-Pois Model: Mixture of Poisson MA(1) Processes

The steps for generating data from the proposed marginalized model using mixture of MA(1) model are same as marginalized model using mixture of AR(1) model except step (6). In this case, the step (6) is: other $T-1$ repeated observations from ith individual, $i =, 1, \cdots, K$ were generated using observation-driven Poisson model as given in either Eq.(4.42) or in Eq.(4.44) conditioning on the value of $d_i = 1$ or $d_i = 0$, respectively. The value of $\rho_1 = \rho_2 = \rho^* = 0.40$ and 0.70 had been considered while using the Eq.(4.42) and the Eq.(4.44).

4.4.3 Marginalized Poisson EQCOR - Poisson EQCOR Mixture Probability Model

Let us consider the stationary equicorrelation based Poisson model (Sutradhar, 2003) for component-1 as

$$y_{it} = \rho_1 * y_{i0} + w_{it}^{(1)}, i = 1, \cdots, K, t = 1, \cdots, T, \tag{4.46}$$

where Y_{i0} ∼Pois$(\mu_{1,i.})$, $W_{it}^{(1)} \overset{\text{iid}}{\sim}$Pois$\big((1-\rho_1)\mu_{1,i.}\big)$ with $\mu_{1,i.} = \exp(\boldsymbol{x}'_{i.}\boldsymbol{\alpha})$, $\boldsymbol{x}'_{i.} = \boldsymbol{x}'_{it}, \forall t = 1, \cdots, T$ and $0 \leq \rho_1 \leq 1$.

Following AR(1) model for component-1, we can derive the expression for mean and variance at all the occasions, and expected value of the products for lag-l. The mean of the ith individual at all the occasions can be computed as

$$E\big[Y_{it}\big] = \mu_{1,i.} = \text{Var}\big[Y_{it}\big], i = 1, \cdots, K, t = 1, \cdots, T.$$

The expected value of the lag-l product of Y_{it} for the ith individual can be computed as

$$E\big[Y_{it}Y_{i,t-l}\big] = \rho_1^2\mu_{1,i.} + \mu_{1,i.}^2, t = 1, \cdots, T, l = 1, \cdots, T-1, t > l.$$

Then the lag-l correlation for component-1, defined as in Eq.(4.2), can be obtained as

$$\rho_l^* = corr\Big[Y_{it}, Y_{i,t-l}\Big] = \rho_1^2. \tag{4.47}$$

One can also shows that the conditional process of Y_{it} ∼Pois$(\mu_{2,i.})$ with Y_{it} in component-2 is defined as

$$y_{it} = \rho_2 * y_{i0} + w_{it}^{(2)}, i = 1, \cdots, K, \tag{4.48}$$

where where $W_{it}^{(2)} \overset{\text{iid}}{\sim}$Pois$\big((1-\rho_2)\mu_{2,i.}\big)$ with $\mu_{2,i.} = \mu_{2,it} \forall t = 1, \cdots, T$, $\mu_{2,i.}$ is defined as in Eq.(4.5) and ρ_2 is defined similarly as ρ_1 for component-1; with the expected value of the joint variable Y_{it} and $Y_{i,t-l}$ as

$$E\Big[Y_{it}Y_{i,t-l}\Big] = \rho_1^2\mu_{1,i.} + \mu_{1,i.}^2.$$

Then the lag-l correlation for component-1, defined as in Eq.(4.2), can be obtained as

$$\rho_l^{**} = corr\Big[Y_{it}, Y_{i,t-l}\Big] = \rho_2^2.$$

The unconditional (marginalized over the subpopulations) mean, and variance and covariance are then computed similarly as Subsection-4.4.2. The unconditional lag-l covariance under Poisson EQCOR - Poisson EQCOR mixture can be obtained as

$$\text{Cov}\Big[Y_{it}, Y_{i,t-l}\Big] = \Big[\pi^*\rho_1^2\mu_{1,i.} + (1-\pi^*)\rho_2^2\mu_{2,i.}\Big] + \Big(\frac{\pi^*}{1-\pi^*}\Big)(\mu_{i.} - \mu_{1,i.})^2.$$

For simplicity, we have considered $\rho_1 = \rho_2 = \rho^*$. Then the covariance between Y_{it} and $Y_{i,t-l}$ becomes

$$\text{Cov}\Big[Y_{it}, Y_{i,t-l}\Big] = \rho^{*2}\mu_{i.} + \Big(\frac{\pi^*}{1-\pi^*}\Big)(\mu_{i.} - \mu_{1,i.})^2.$$

Therefore, the unconditional lag-l correlation defined in Eq.(4.7) for the ith individual under marginalized Poisson EQCOR-Poisson EQCOR mixture model setup can be obtained as

$$\rho_{l,i} = \text{corr}\Big[Y_{it}, Y_{i,t-l}\Big] = \frac{\rho^{*2}\mu_{i.} + \Big(\frac{\pi^*}{1-\pi^*}\Big)(\mu_{i.} - \mu_{1,i.})^2}{\mu_{i.} + \big(\frac{\pi^*}{1-\pi^*}\big)(\mu_{i.} - \mu_{1,i.})^2}. \tag{4.49}$$

Note that Eq.(4.49) reduces to correlation structure of Poisson EQCOR probability model (Sutradhar, 2003) if and only if $\mu_{1,i.} = \mu_{2,i.} = \mu_{i.}$. Finally, the longitudinal correlation under proposed marginalized model is obtained by averaging the K quantities of Eq.(4.49) using Eq.(4.41).

Data Generation from RMMPois-Pois Model: Mixture of Poisson EQCOR Processes

The steps for generating data from the proposed marginalized model using mixture of EQCOR model are same as marginalized model using mixture of AR(1) model except step (6). In this case, the step (6) is: other $T-1$ repeated observations from ith individual, $i\ =, 1, \cdots, K$ were generated using observation-driven Poisson model as given in either

Eq.(4.46) or in Eq.(4.48) conditioning on the value of $d_i = 1$ or $d_i = 0$, respectively. The value of $\rho_1 = \rho_2 = \rho^* = 0.40$ and 0.70 had been considered while using the Eq.(4.46) and the Eq.(4.48).

4.4.4 Performance of the Proposed RMMPois-Pois Model

The simulation was repeated 1000 times for each setup. We estimated the regression parameters (modeling marginalized means), the lag-l correlation parameters, all other nuisance parameters by the proposed estimation approaches which are given in Section-4.3. In order to investigate the performance of the estimates, we have computed the bias, standard error and the coverage probability (Cov.Pr.). The biases were computed by taking the differences between simulated means (SM) and the true values for each of the parameters. Two types of standard errors such as estimated standard errors (ESE) and simulated standard errors (SSE) were also computed to investigate the properties of the estimators. The SM, ESE and SSE for estimators $\hat{\beta}$ and $\hat{\alpha}$ were calculated similarly as in Eq.(2.29). The proportion of convergences (Conv.Prop.) in fitting the REMPois-Pois model were computed for all the setups.

The SM, Bias, ESE, SSE, Cov.Pr. and Conv.Prop. were computed from RMMPois-Pois model using mixture of Poisson AR(1), mixture of Poisson MA(1) and mixture of Poisson EQCOR model. The results obtained from mixing proportion $\pi^* = 0.70$ with $\rho^* = 0.40$ were given in Table-4.1, Table-4.2, and Table-4.3 for mixture of Poisson AR(1), mixture of Poisson MA(1) and mixture of Poisson EQCOR model, respectively. The results of the simulation studies were also given in the Appendix C for mixing proportion $\pi^* = 0.50$ with $\rho^* = 0.40$ in Table-C.1, Table-C.3, and Table-C.5 for mixture of Poisson AR(1), mixture of Poisson MA(1) and mixture of Poisson EQCOR model respectively; and for mixing proportion

$\pi^* = 0.90$ with $\rho^* = 0.40$ in Table-C.2, Table-C.4, and Table-C.6 for mixture of Poisson AR(1), mixture of Poisson MA(1) and mixture of Poisson EQCOR model, respectively.

Table 4.1: Simulated mean (SM), Bias, estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.70$ from marginalized mixture of Poisson AR(1) model for different values of K and T

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(100,3)		$\beta_0 = 0.20$	0.196	-0.004	0.292	0.259	96.0
		$\beta_1 = 0.60$	0.570	-0.030	0.451	0.424	95.2
		$\beta_2 = 0.50$	0.491	-0.009	0.317	0.284	95.5
		$\rho_1 = 0.810$	0.805	-0.005			
		$\alpha_0 = -1.00$	-0.942	0.058	0.301	0.411	95.1
		$\alpha_1 = 0.40$	0.325	-0.075	0.452	0.591	94.7
	0.871	$\alpha_2 = 0.50$	0.455	-0.045	0.308	0.394	94.8
		$\rho_1^* = 0.40$	0.337	-0.063			
		$\pi^* = 0.70$	0.694	-0.006			
(100,4)		$\beta_0 = 0.20$	0.195	-0.005	0.286	0.238	94.8
		$\beta_1 = 0.60$	0.592	-0.008	0.442	0.388	94.8
		$\beta_2 = 0.50$	0.478	-0.022	0.310	0.283	95.2
		$\rho_1 = 0.810$	0.806	-0.004			
		$\alpha_0 = -1.00$	-0.949	0.051	0.275	0.366	93.9
		$\alpha_1 = 0.40$	0.345	-0.055	0.413	0.512	94.3
	0.884	$\alpha_2 = 0.50$	0.463	-0.037	0.280	0.357	95.0
		$\rho_1^* = 0.40$	0.367	-0.033			
		$\pi^* = 0.70$	0.694	-0.006			
(100,5)		$\beta_0 = 0.20$	0.195	-0.005	0.284	0.256	95.7
		$\beta_1 = 0.60$	0.575	-0.025	0.438	0.409	94.8
		$\beta_2 = 0.50$	0.489	-0.011	0.308	0.270	95.6
		$\rho_1 = 0.810$	0.807	-0.003			
		$\alpha_0 = -1.00$	-0.946	0.054	0.255	0.326	94.6
		$\alpha_1 = 0.40$	0.351	-0.049	0.382	0.459	94.4
	0.881	$\alpha_2 = 0.50$	0.460	-0.040	0.260	0.307	94.0
		$\rho_1^* = 0.40$	0.386	-0.014			
		$\pi^* = 0.70$	0.695	-0.005			
(200,3)		$\beta_0 = 0.20$	0.205	0.005	0.208	0.184	95.9
		$\beta_1 = 0.60$	0.584	-0.016	0.324	0.291	95.7
		$\beta_2 = 0.50$	0.490	-0.010	0.211	0.185	95.5
		$\rho_1 = 0.810$	0.811	0.001			
		$\alpha_0 = -1.00$	-0.934	0.066	0.216	0.291	95.5
		$\alpha_1 = 0.40$	0.343	-0.057	0.326	0.413	94.6
	0.905	$\alpha_2 = 0.50$	0.455	-0.045	0.205	0.258	93.9
		$\rho_1^* = 0.40$	0.353	-0.047			
		$\pi^* = 0.70$	0.696	-0.004			
(200,4)		$\beta_0 = 0.20$	0.173	-0.027	0.204	0.180	95.7
		$\beta_1 = 0.60$	0.620	0.020	0.320	0.286	95.7
		$\beta_2 = 0.50$	0.518	0.018	0.208	0.197	95.7
		$\rho_1 = 0.810$	0.812	0.002			

Continued...Table 4.1

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
		$\alpha_0 = -1.00$	-0.925	0.075	0.195	0.259	94.5
		$\alpha_1 = 0.40$	0.320	-0.080	0.297	0.373	94.5
	0.913	$\alpha_2 = 0.50$	0.457	-0.043	0.186	0.222	94.6
		$\rho_1^* = 0.40$	0.373	-0.027			
		$\pi^* = 0.70$	0.698	-0.002			
(200,5)		$\beta_0 = 0.20$	0.190	-0.010	0.202	0.163	94.7
		$\beta_1 = 0.60$	0.602	0.002	0.315	0.278	96.4
		$\beta_2 = 0.50$	0.496	-0.004	0.205	0.175	94.7
		$\rho_1 = 0.810$	0.812	0.002			
		$\alpha_0 = -1.00$	-0.915	0.085	0.181	0.225	93.0
		$\alpha_1 = 0.40$	0.312	-0.088	0.275	0.322	93.7
	0.867	$\alpha_2 = 0.50$	0.441	-0.059	0.172	0.202	94.1
		$\rho_1^* = 0.40$	0.390	-0.010			
		$\pi^* = 0.70$	0.698	-0.002			
(500,3)		$\beta_0 = 0.20$	0.200	0.000	0.125	0.111	95.0
		$\beta_1 = 0.60$	0.588	-0.012	0.212	0.191	95.5
		$\beta_2 = 0.50$	0.502	0.002	0.137	0.121	96.0
		$\rho_1 = 0.810$	0.809	-0.001			
		$\alpha_0 = -1.00$	-0.942	0.058	0.129	0.166	94.5
		$\alpha_1 = 0.40$	0.366	-0.034	0.211	0.253	93.9
	0.902	$\alpha_2 = 0.50$	0.466	-0.034	0.130	0.157	95.3
		$\rho_1^* = 0.40$	0.356	-0.044			
		$\pi^* = 0.70$	0.699	-0.001			
(500,4)		$\beta_0 = 0.20$	0.207	0.007	0.123	0.103	95.5
		$\beta_1 = 0.60$	0.587	-0.013	0.208	0.181	96.1
		$\beta_2 = 0.50$	0.487	-0.013	0.134	0.122	95.7
		$\rho_1 = 0.810$	0.809	-0.001			
		$\alpha_0 = -1.00$	-0.910	0.090	0.118	0.145	89.5
		$\alpha_1 = 0.40$	0.314	-0.086	0.193	0.226	93.0
	0.898	$\alpha_2 = 0.50$	0.441	-0.059	0.119	0.138	93.0
		$\rho_1^* = 0.40$	0.382	-0.018			
		$\pi^* = 0.70$	0.699	-0.001			
(500,5)		$\beta_0 = 0.20$	0.199	-0.001	0.121	0.099	94.9
		$\beta_1 = 0.60$	0.595	-0.005	0.206	0.173	95.7
		$\beta_2 = 0.50$	0.498	-0.002	0.133	0.113	95.8
		$\rho_1 = 0.810$	0.810	-0.000			
		$\alpha_0 = -1.00$	-0.922	0.078	0.110	0.130	91.1
		$\alpha_1 = 0.40$	0.321	-0.079	0.180	0.202	93.8
	0.890	$\alpha_2 = 0.50$	0.450	-0.050	0.111	0.121	93.1
		$\rho_1^* = 0.40$	0.398	-0.002			
		$\pi^* = 0.70$	0.699	-0.001			

Table 4.2: Simulated mean (SM), Bias, estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.70$ from marginalized mixture of Poisson MA(1) model for different values of K and T

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(100,3)		$\beta_0 = 0.20$	0.209	0.009	0.286	0.249	95.1
		$\beta_1 = 0.60$	0.555	-0.045	0.443	0.400	95.1
		$\beta_2 = 0.50$	0.464	-0.036	0.311	0.287	94.9
		$\rho_1 = 0.775$	0.770	-0.005			
		$\alpha_0 = -1.00$	-0.919	0.081	0.279	0.370	93.8
		$\alpha_1 = 0.40$	0.318	-0.082	0.420	0.544	94.4
	0.896	$\alpha_2 = 0.50$	0.437	-0.063	0.286	0.360	94.0
		$\rho_1^* = 0.286$	0.256	-0.030			
		$\pi^* = 0.70$	0.697	-0.003			
(100,4)		$\beta_0 = 0.20$	0.185	-0.015	0.283	0.250	95.7
		$\beta_1 = 0.60$	0.602	0.002	0.437	0.405	95.5
		$\beta_2 = 0.50$	0.484	-0.016	0.307	0.277	95.3
		$\rho_1 = 0.775$	0.771	-0.004			
		$\alpha_0 = -1.00$	-0.936	0.064	0.250	0.310	94.9
		$\alpha_1 = 0.40$	0.341	-0.059	0.375	0.426	94.2
	0.890	$\alpha_2 = 0.50$	0.441	-0.059	0.256	0.301	94.2
		$\rho_1^* = 0.286$	0.275	-0.011			
		$\pi^* = 0.70$	0.696	-0.004			
(100,5)		$\beta_0 = 0.20$	0.193	-0.007	0.280	0.256	95.9
		$\beta_1 = 0.60$	0.574	-0.026	0.434	0.408	96.1
		$\beta_2 = 0.50$	0.481	-0.019	0.304	0.276	96.4
		$\rho_1 = 0.775$	0.772	-0.003			
		$\alpha_0 = -1.00$	-0.926	0.074	0.228	0.280	93.1
		$\alpha_1 = 0.40$	0.311	-0.089	0.343	0.406	94.4
	0.899	$\alpha_2 = 0.50$	0.449	-0.051	0.232	0.269	92.5
		$\rho_1^* = 0.286$	0.288	0.003			
		$\pi^* = 0.70$	0.699	-0.001			
(200,3)		$\beta_0 = 0.20$	0.192	-0.008	0.204	0.173	95.3
		$\beta_1 = 0.60$	0.598	-0.002	0.319	0.276	95.7
		$\beta_2 = 0.50$	0.502	0.002	0.208	0.188	95.5
		$\rho_1 = 0.775$	0.776	0.001			
		$\alpha_0 = -1.00$	-0.917	0.083	0.200	0.258	93.5
		$\alpha_1 = 0.40$	0.331	-0.069	0.302	0.373	93.7
	0.893	$\alpha_2 = 0.50$	0.439	-0.061	0.190	0.232	94.4
		$\rho_1^* = 0.286$	0.265	-0.020			
		$\pi^* = 0.70$	0.697	-0.003			
(200,4)		$\beta_0 = 0.20$	0.195	-0.005	0.202	0.171	95.5
		$\beta_1 = 0.60$	0.584	-0.016	0.316	0.278	95.3
		$\beta_2 = 0.50$	0.497	-0.003	0.206	0.183	96.5
		$\rho_1 = 0.775$	0.776	0.001			
		$\alpha_0 = -1.00$	-0.934	0.066	0.179	0.231	93.9
		$\alpha_1 = 0.40$	0.332	-0.068	0.272	0.332	94.9
	0.887	$\alpha_2 = 0.50$	0.457	-0.043	0.170	0.196	95.5
		$\rho_1^* = 0.286$	0.279	-0.007			
		$\pi^* = 0.70$	0.700	-0.000			

Continued...Table 4.2

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(200,5)		$\beta_0 = 0.20$	0.195	-0.005	0.199	0.171	95.8
		$\beta_1 = 0.60$	0.596	-0.004	0.311	0.283	95.1
		$\beta_2 = 0.50$	0.499	-0.001	0.202	0.182	95.6
		$\rho_1 = 0.775$	0.777	0.002			
		$\alpha_0 = -1.00$	-0.927	0.073	0.164	0.200	93.5
		$\alpha_1 = 0.40$	0.324	-0.076	0.249	0.279	93.9
	0.903	$\alpha_2 = 0.50$	0.450	-0.050	0.156	0.178	93.8
		$\rho_1^* = 0.286$	0.293	0.007			
		$\pi^* = 0.70$	0.697	-0.003			
(500,3)		$\beta_0 = 0.20$	0.205	0.005	0.123	0.103	96.2
		$\beta_1 = 0.60$	0.584	-0.016	0.209	0.180	95.3
		$\beta_2 = 0.50$	0.488	-0.012	0.135	0.123	95.1
		$\rho_1 = 0.775$	0.773	-0.002			
		$\alpha_0 = -1.00$	-0.915	0.085	0.120	0.157	91.9
		$\alpha_1 = 0.40$	0.333	-0.067	0.196	0.235	93.7
	0.890	$\alpha_2 = 0.50$	0.449	-0.051	0.121	0.153	93.5
		$\rho_1^* = 0.286$	0.267	-0.019			
		$\pi^* = 0.70$	0.700	0.000			
(500,4)		$\beta_0 = 0.20$	0.195	-0.005	0.121	0.100	95.5
		$\beta_1 = 0.60$	0.601	0.001	0.205	0.174	96.1
		$\beta_2 = 0.50$	0.500	-0.000	0.132	0.118	95.5
		$\rho_1 = 0.775$	0.773	-0.002			
		$\alpha_0 = -1.00$	-0.925	0.075	0.108	0.129	91.7
		$\alpha_1 = 0.40$	0.325	-0.075	0.177	0.196	93.1
	0.882	$\alpha_2 = 0.50$	0.453	-0.047	0.109	0.127	93.8
		$\rho_1^* = 0.286$	0.288	0.002			
		$\pi^* = 0.70$	0.699	-0.001			
(500,5)		$\beta_0 = 0.20$	0.193	-0.007	0.120	0.098	95.5
		$\beta_1 = 0.60$	0.612	0.012	0.203	0.180	96.0
		$\beta_2 = 0.50$	0.500	-0.000	0.131	0.122	95.8
		$\rho_1 = 0.775$	0.773	-0.002			
		$\alpha_0 = -1.00$	-0.937	0.063	0.099	0.113	90.0
		$\alpha_1 = 0.40$	0.336	-0.064	0.162	0.175	93.3
	0.929	$\alpha_2 = 0.50$	0.461	-0.039	0.100	0.109	92.8
		$\rho_1^* = 0.286$	0.299	0.013			
		$\pi^* = 0.70$	0.698	-0.002			

Table 4.3: Simulated mean (SM), Bias, estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.70$ from marginalized mixture of Poisson EQCOR model for different values of K and T

(K, T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(100,3)		$\beta_0 = 0.20$	0.189	-0.011	0.287	0.269	95.8
		$\beta_1 = 0.60$	0.578	-0.022	0.444	0.419	95.6
		$\beta_2 = 0.50$	0.489	-0.011	0.313	0.308	95.2
		$\rho_1 = 0.735$	0.733	-0.002			
		$\alpha_0 = -1.00$	-0.927	0.073	0.275	0.383	93.5
		$\alpha_1 = 0.40$	0.312	-0.088	0.414	0.544	94.0
	0.902	$\alpha_2 = 0.50$	0.430	-0.070	0.283	0.369	94.7
		$\rho_1^* = 0.16$	0.150	-0.010			
		$\pi^* = 0.70$	0.696	-0.004			
(100,4)		$\beta_0 = 0.20$	0.189	-0.011	0.285	0.247	96.4
		$\beta_1 = 0.60$	0.592	-0.008	0.440	0.397	96.1
		$\beta_2 = 0.50$	0.488	-0.012	0.310	0.281	96.1
		$\rho_1 = 0.735$	0.731	-0.004			
		$\alpha_0 = -1.00$	-0.935	0.065	0.254	0.318	95.1
		$\alpha_1 = 0.40$	0.323	-0.077	0.382	0.459	95.9
	0.900	$\alpha_2 = 0.50$	0.440	-0.060	0.261	0.312	94.1
		$\rho_1^* = 0.16$	0.155	-0.005			
		$\pi^* = 0.70$	0.697	-0.003			
(100,5)		$\beta_0 = 0.20$	0.185	-0.015	0.282	0.264	95.6
		$\beta_1 = 0.60$	0.592	-0.008	0.436	0.413	95.9
		$\beta_2 = 0.50$	0.500	-0.000	0.306	0.283	95.6
		$\rho_1 = 0.735$	0.732	-0.003			
		$\alpha_0 = -1.00$	-0.949	0.051	0.241	0.307	93.7
		$\alpha_1 = 0.40$	0.338	-0.062	0.362	0.431	95.1
	0.906	$\alpha_2 = 0.50$	0.469	-0.031	0.246	0.307	94.6
		$\rho_1^* = 0.16$	0.159	-0.001			
		$\pi^* = 0.70$	0.696	-0.004			
(200,3)		$\beta_0 = 0.20$	0.189	-0.011	0.205	0.186	95.9
		$\beta_1 = 0.60$	0.604	0.004	0.320	0.294	94.8
		$\beta_2 = 0.50$	0.498	-0.002	0.208	0.195	95.7
		$\rho_1 = 0.735$	0.738	0.003			
		$\alpha_0 = -1.00$	-0.935	0.065	0.198	0.247	94.6
		$\alpha_1 = 0.40$	0.358	-0.042	0.299	0.345	94.9
	0.922	$\alpha_2 = 0.50$	0.454	-0.046	0.188	0.225	94.8
		$\rho_1^* = 0.16$	0.161	0.001			
		$\pi^* = 0.70$	0.698	-0.002			
(200,4)		$\beta_0 = 0.20$	0.193	-0.007	0.202	0.162	94.9
		$\beta_1 = 0.60$	0.596	-0.004	0.317	0.274	94.7
		$\beta_2 = 0.50$	0.501	0.001	0.206	0.177	95.6
		$\rho_1 = 0.735$	0.739	0.004			
		$\alpha_0 = -1.00$	-0.931	0.069	0.182	0.232	94.8
		$\alpha_1 = 0.40$	0.327	-0.073	0.277	0.323	94.9
	0.885	$\alpha_2 = 0.50$	0.453	-0.047	0.173	0.207	94.6
		$\rho_1^* = 0.16$	0.164	0.004			
		$\pi^* = 0.70$	0.698	-0.002			

Continued...Table 4.3

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(200,5)		$\beta_0 = 0.20$	0.200	-0.000	0.201	0.172	95.1
		$\beta_1 = 0.60$	0.594	-0.006	0.314	0.279	96.4
		$\beta_2 = 0.50$	0.485	-0.015	0.204	0.192	95.9
		$\rho_1 = 0.735$	0.738	0.003			
		$\alpha_0 = -1.00$	-0.935	0.065	0.173	0.219	93.3
		$\alpha_1 = 0.40$	0.332	-0.068	0.262	0.311	93.6
	0.927	$\alpha_2 = 0.50$	0.452	-0.048	0.164	0.200	93.3
		$\rho_1^* = 0.16$	0.168	0.008			
		$\pi^* = 0.70$	0.697	-0.003			
(500,3)		$\beta_0 = 0.20$	0.199	-0.001	0.123	0.102	95.2
		$\beta_1 = 0.60$	0.597	-0.003	0.208	0.175	95.2
		$\beta_2 = 0.50$	0.499	-0.001	0.134	0.119	96.1
		$\rho_1 = 0.735$	0.734	-0.001			
		$\alpha_0 = -1.00$	-0.917	0.083	0.118	0.151	91.5
		$\alpha_1 = 0.40$	0.334	-0.066	0.194	0.231	94.8
	0.893	$\alpha_2 = 0.50$	0.445	-0.055	0.119	0.142	93.6
		$\rho_1^* = 0.16$	0.165	0.005			
		$\pi^* = 0.70$	0.699	-0.001			
(500,4)		$\beta_0 = 0.20$	0.193	-0.007	0.122	0.101	96.0
		$\beta_1 = 0.60$	0.602	0.002	0.206	0.173	95.8
		$\beta_2 = 0.50$	0.502	0.002	0.133	0.115	96.1
		$\rho_1 = 0.735$	0.733	-0.002			
		$\alpha_0 = -1.00$	-0.932	0.068	0.110	0.132	92.6
		$\alpha_1 = 0.40$	0.330	-0.070	0.181	0.206	93.3
	0.890	$\alpha_2 = 0.50$	0.453	-0.047	0.111	0.125	94.7
		$\rho_1^* = 0.16$	0.173	0.013			
		$\pi^* = 0.70$	0.699	-0.001			
(500,5)		$\beta_0 = 0.20$	0.202	0.002	0.121	0.102	95.6
		$\beta_1 = 0.60$	0.588	-0.012	0.205	0.175	96.0
		$\beta_2 = 0.50$	0.499	-0.001	0.132	0.120	96.0
		$\rho_1 = 0.735$	0.733	-0.002			
		$\alpha_0 = -1.00$	-0.934	0.066	0.104	0.131	91.9
		$\alpha_1 = 0.40$	0.331	-0.069	0.171	0.203	93.6
	0.890	$\alpha_2 = 0.50$	0.453	-0.047	0.105	0.121	93.5
		$\rho_1^* = 0.16$	0.175	0.015			
		$\pi^* = 0.70$	0.700	-0.000			

From Table-4.1-Table-4.3, it is clear that the estimators of marginal parameters and correlation parameters had minimal amount of biases for all the settings. For example, the biases in Table-4.1 for mixture of Poisson AR(1) model are given as follows. The amount of bias of $(\beta_0, \beta_1, \beta_2)$ was $(-0.004, -0.030, -0.009)$ when $T = 3$ and it was $(-0.005, -0.025, -0.011)$ when $T = 5$ for $K = 100$. For $K = 500$, these were (0.000,-

0.012,0.002) and $(-0.001, -0.005, -0.002)$ for $T = 3$ and $T = 5$, respectively. The amount of bias for lag-1 correlation of marginalized model (ρ_1) was -0.005 when $T = 3$ and -0.003 when $T = 5$ for $K = 100$. These amount were -0.001 and 0.000 for $T = 3$ and $T = 5$, respectively when we considered $K = 500$. Although the amount of biases are similar for different values of T, these amount decreases with increasing the values of K for most of the parameters. Similar pattern had also been observed for the nuisance parameters. The rate of convergence increases with increasing the value of K. It was found that the coverage probabilities are almost equal to the nominal level of confidence for all the parameters. The ESE and SSE were almost same for all the parameters. It was also observed that the results of Table-4.2, and Table-4.3 follow the similar pattern of the results given in Table-4.1.

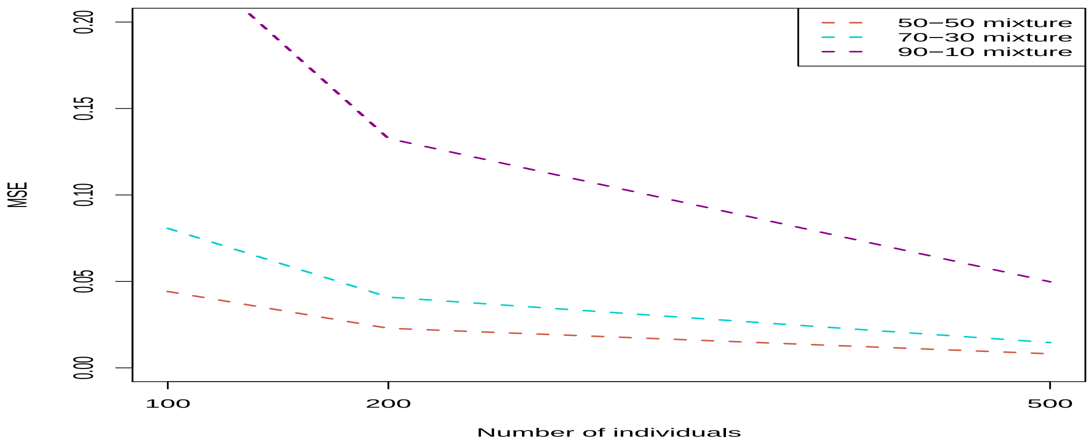
For mixing proportion $\pi^* = 0.50$, the results of simulation studies from mixture of AR(1), mixture of MA(1) and mixture of EQCOR were given in Table-C.1, Table-C.3, and Table-C.5, respectively. From the results, it had been observed that the estimates of marginal parameters and other nuisance parameters had a minimal amount of bias for all settings. The relationship between bias and the values of K, and T was found to be similar to that observed for $\pi^* = 0.70$.

For mixing proportion $\pi^* = 0.90$, the SM, Bias, ESE, SSE, Cov.Pr. and Conv.Prop. computed from the simulation studies for mixture of AR(1), mixture of MA(1) and mixture of EQCOR were given in Table-C.2, Table-C.4, and Table-C.6, respectively. The pattern of the results were found to be similar to the results observed for $\pi^* = 0.50$ and $\pi^* = 0.90$ in estimating all the marginal parameters and other nuisance parameters. However, the biases obtained for $\pi^* = 0.90$ were found to be higher than the biases obtained for mixing proportion $\pi^* = 0.50$ and $\pi^* = 0.70$.

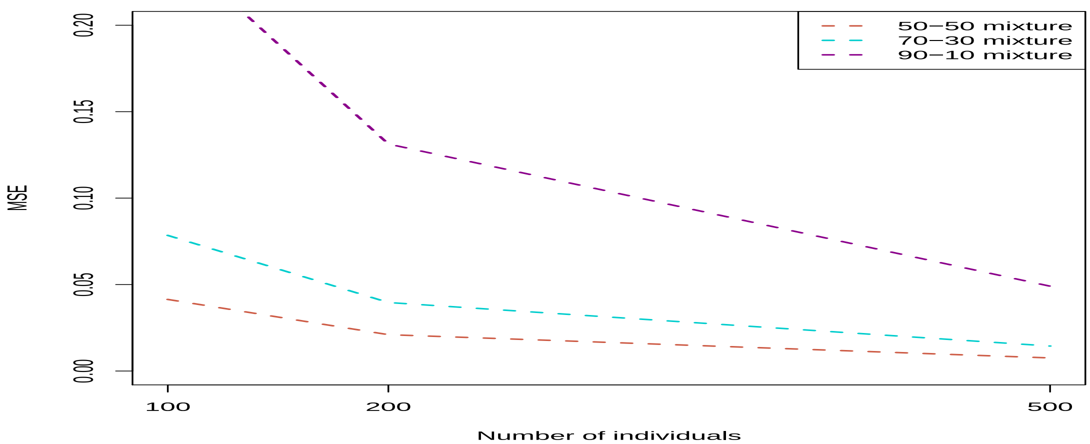
The MSE of $\beta_0, \beta_1, \beta_2$ for various observation driven probability model and for different values of mixing probabilities were shown in Figure-4.1–Figure-4.3, respectively. In each Figure, part (a), part (b) and part (c) were drawn for mixture of AR(1), mixture of MA(1) and mixture of EQCOR models, respectively. From these figures, it is depicted that the MSEs were found low for all the indicated situations except for $\pi^* = 0.90$. Also, the MSEs decreases for increasing the value of K.

The Bias for the nuisance parameter ρ_1 (lag-1 correlation) for different observation driven model and for different values of mixing probabilities was presented in Figure-4.4. It also contains part (a), part (b) and part (c) for representing Bias in case of mixture of AR(1), mixture of MA(1) and mixture of EQCOR models, respectively. From the figure, it is clear that the Bias were found close to zero for all the indicated situations except for $\pi^* = 0.90$ and $K = 100$.

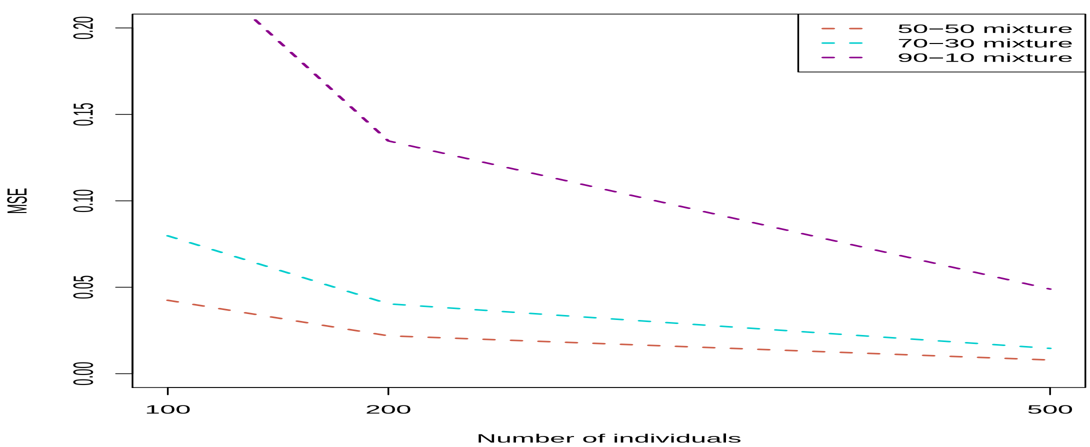
Extensive simulation studies were also conducted for $\rho^* = 0.70$. Results for simulation study obtained using $\rho^* = 0.70$ from mixture of Poisson AR(1), Poisson MA(1), and Poisson EQCOR models with mixing proportion $\pi^* = 0.50$, $\pi^* = 0.70$ and $\pi^* = 0.90$ were given in Table-C.7, Table-C.8 and Table-C.9, respectively in the Appendix C (Section C.4) for $T = 4$ and for different values of K.



(a) Number of individuals vs. MSE for mixture of Poisson AR(1)

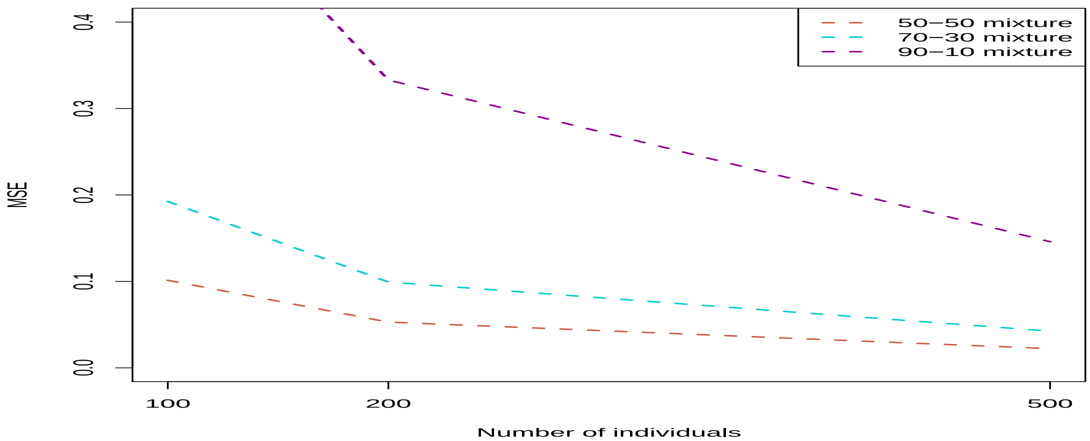


(b) Number of individuals vs. MSE mixture of Poisson MA(1)

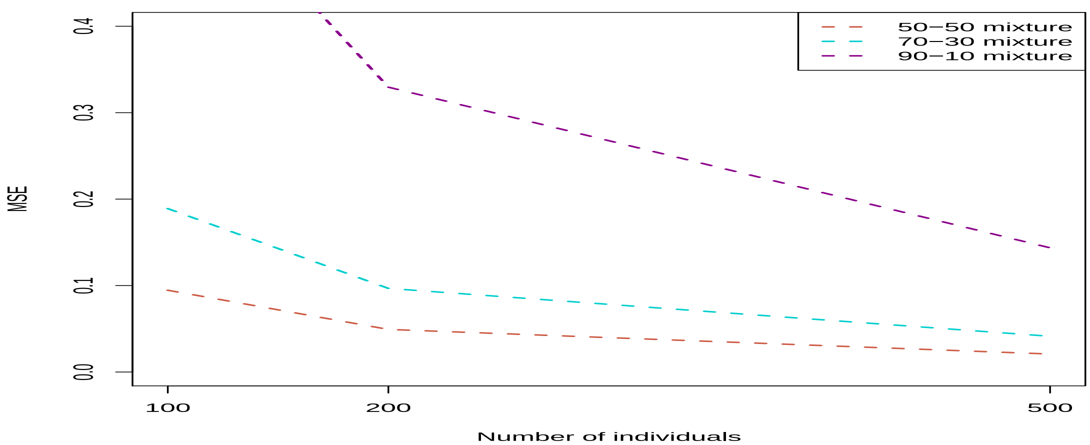


(c) Number of individuals vs. MSE mixture of Poisson EQCOR

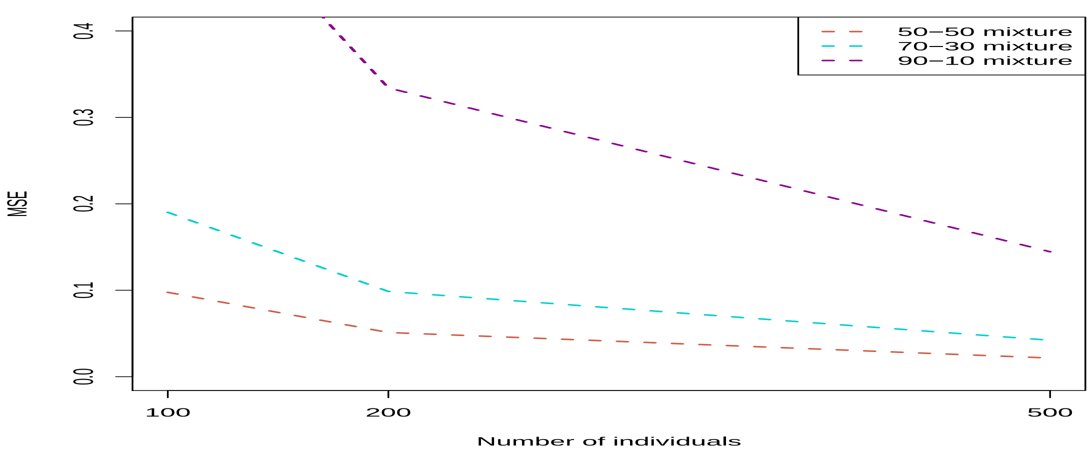
Figure 4.1: A comparison of MSE from RMMPois-Pois models with different mixing probability and varying number of individuals for the regression parameter $\beta_0 = 0.20$



(a) Number of individuals vs. MSE for mixture of Poisson AR(1)

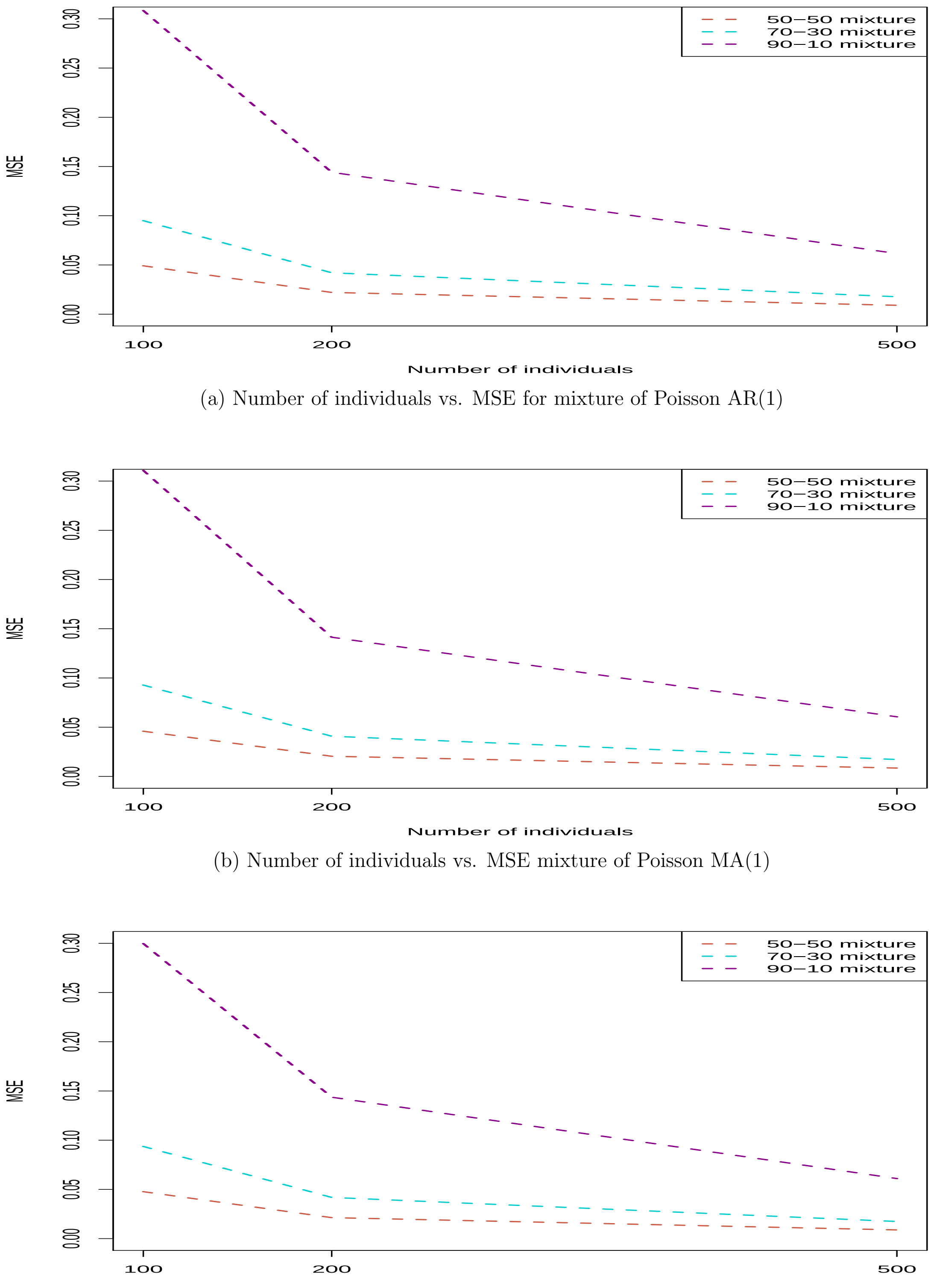


(b) Number of individuals vs. MSE mixture of Poisson MA(1)



(c) Number of individuals vs. MSE mixture of Poisson EQCOR

Figure 4.2: A comparison of MSE from RMMPois-Pois models with different mixing probability and varying number of individuals for the regression parameter $\beta_1 = 0.60$

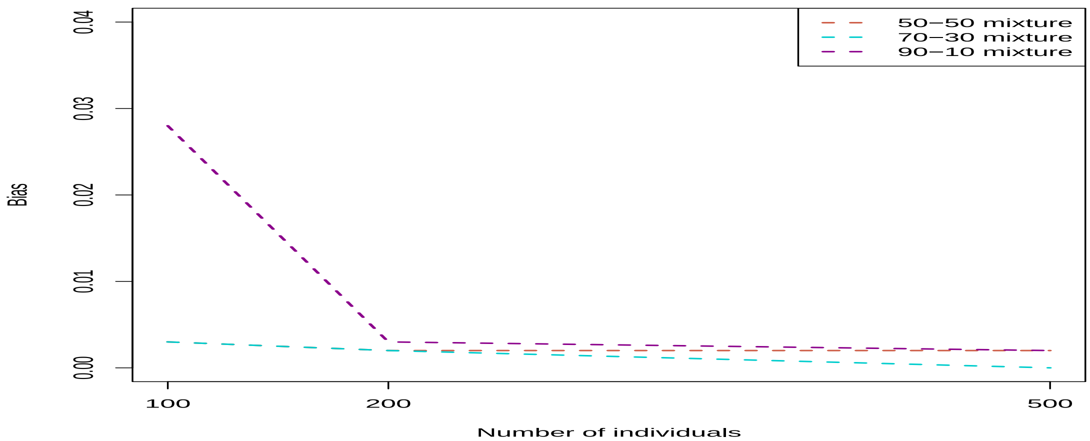


(a) Number of individuals vs. MSE for mixture of Poisson AR(1)

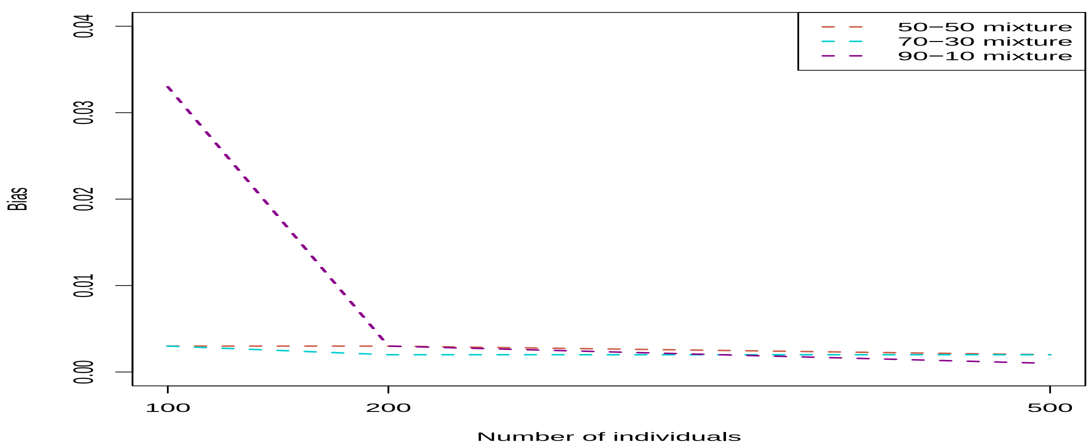
(b) Number of individuals vs. MSE mixture of Poisson MA(1)

(c) Number of individuals vs. MSE mixture of Poisson EQCOR

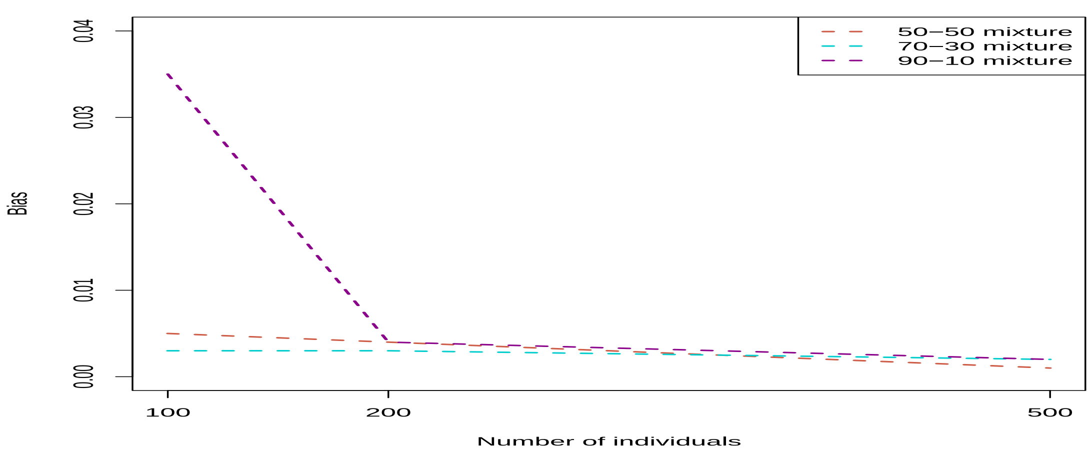
Figure 4.3: A comparison of MSE from RMMPois-Pois models with different mixing probability and varying number of individuals for the regression parameter $\beta_2 = 0.50$



(a) Number of individuals vs. Bias for mixture of Poisson AR(1)



(b) Number of individuals vs. Bias mixture of Poisson MA(1)



(c) Number of individuals vs. Bias mixture of Poisson EQCOR

Figure 4.4: A comparison of Bias from RMMPois-Pois models with different mixing probability and varying number of individuals for the lag-1 correlation ρ_1

4.5 Illustration

Min and Agresti (2005) had conducted a study for analyzing zero-inflated repeated measures data, where the data set was obtained from a pharmaceutical company. The main aspect of the study conducted by Min and Agresti (2005) was to make a comparison between two treatments for a particular disease using repeated number of episodes of a certain side effect. To illustrate our proposed (RMMPois-Pois) model to analyze the longitudinal zero-inflated count data, we have utilized the same data set and the data set is available on the journal website https://journals.sagepub.com/doi/abs/10.1191/1471082X05st084oa. In the data set 118 patients were considered. Of them, 59 patients were randomly allocated to treatment A (TRT1) and rest of them were allocated to treatment B (TRT2). The number of side effect episodes was measured at each of six visits (i.e., $T = 6$). The data set contains observations about 83% zero counts. Following the notations used in the proposed model, the covariate (type of treatment) for the ith individual ($i = 1, \cdots, K = 118$) at time t ($t = 1, \cdots, 6$) is denoted by x_{it} . Let x_{it} be the value of covariate for the ith individual at tth time point. Since the covariate is time independent, $x_{it} = x_{i.} \forall t = 1, \cdots, T$. We have $x_{i.} = 0$, if the ith individual received treatment A and $x_{it} = 1$, if the individual received treatment B. The effect of the covariate is denoted by β_1 in the population and α_1 in the component-1 so that the marginal mean (marginalized over subpopulation) and the latent component-1 mean of the count responses for the ith individual is given by

$$\mu_{i.} = \exp(\beta_0 + \beta_1 x_{i.}); \quad \text{and} \quad \mu_{1,i.} = \exp(\alpha_0 + \alpha_1 x_{i.}), \tag{4.50}$$

where β_0 and α_0 are intercept terms corresponding to the marginalized model and the component-1 model, respectively.

If we assume the repeated responses $y_{i1}, \cdots, y_{i6}$ are correlated with true correlation structure $C(\rho)$ in the population and $C_1(\rho^*)$ in the latent component-1, then the GQL based regression estimates and the moment estimates of true correlations can be obtained by following the estimation approach discussed in Section-4.3. The estimates of regression parameters, along with the lag-1 correlation, standard errors (SE), and IRR obtained from RMMPois-Pois model are given in Table-4.4.

Table 4.4: RMMPois-Pois Model estimates of parameters, standard errors (SE), and the IRR for analyzing repeated number of episodes of a certain side effect

Variable name	Estimates	SE	p-value	IRR
Marginalized:				
Intercept	-1.871	0.200	<0.001	
Treatment				
TRT1 (ref)				
TRT2	0.984	0.239	<0.001	2.675
Lag-1 correlation				
ρ_1	0.427			
Component-1:				
Intercept	-2.094	0.194	<0.001	
Treatment				
TRT1 (ref)				
TRT2	0.887	0.232	<0.001	-
Lag-1 correlation				
ρ_1^*	0.338			

It was observed from Table-4.4 that the treatment had significant (p-value < 0.001) effect on the number of episode of side effect. The incidence rate of side effect for the patient taking Treatment B was 2.675 times of the incidence rate of treatment A. The estimate of true lag-1 correlation was found to be 0.427 in the overall population. Estimates of other lag correlations are presented in Table-4.5.

Table 4.5: Lag-l correlation estimated from the proposed RMMPois-Pois model

l	2	3	4	5
ρ_l	0.273	0.116	0.047	0.012

4.6 Conclusion

In this chapter, we have assumed that the zero-inflated repeated count data arise from two latent populations and hence we proposed a repeated measures Poisson model for each of the component of mixture. Therefore, the proposed approach allowed us to control the unexplained heterogeneity accounted for by latent classes and also to address the excess zeros in the analysis. In addition, the suggested model can directly postulate the marginal mean to draw straightforward inference for overall exposure effects taking the correlation among the repeated measures of an individual into account.

The performance of the proposed model has been examined by conducting extensive simulation studies for mixture of observation-driven Poisson model such as Poisson AR(1), Poisson MA(1) and Poisson equicorrelation models. It was observed from the simulation studies that the RMMPois-Pois model, in general, provides estimates with low bias and low MSE.

In this study, a data set from a pharmaceutical company was analyzed by using the proposed model as an illustration in order to make a comparison between two treatments for a particular disease in terms of the repeated number of episodes for certain side effect.

Chapter 5

Conclusions

Mixture models control heterogeneity in data by decomposing the population into two or more latent subgroups each of which is modeled by using its own set of parameters (McLachlan and Peel, 2000; Benecha et al., 2017). Although mixture of a 'not-at-risk' class and 'at-risk' class are usually used for analyzing zero-inflated data in the literature (Lambert, 1992; Hall, 2000; Hasan and Sneddon, 2009; Long et al., 2014, 2015), mixture of two 'at-risk' classes can also be used to analyze zero-inflated count data when all individuals have a risk of developing the event of interest (Benecha et al., 2017; Wang et al., 2007). This is because if the population consists mixture of 'at-risk' and 'not-at-risk' groups for analyzing zero-inflated Poisson data, a zero-truncated Poisson model provide similar inference as zero-inflated Poisson model (Haque et al., 2022b).

In order to model the marginal means for analyzing heterogeneous count data, two-component finite mixture distributions have been used by researchers (Long et al., 2015; Benecha et al., 2017; Long et al., 2014). In such cases, the model is built in such a way that it can directly estimate the regression parameters for the marginal mean.

Although non-standard and standard two-component mixture models have been developed under correlated count responses (Wang et al., 2007, 2002; Hasan and Sneddon, 2009), no researcher has yet developed the marginalized version of such a model to the best of our knowledge. This motivates us to develop a marginalized Poisson-Poisson mixture model (MPois-Pois) for analysing zero-inflated correlated count data.

As an extension of the Poisson-Poisson mixture model (Wang et al., 1996) under cross-sectional data contexts, Benecha et al. (2017) have proposed a marginalized version of the Poisson-Poisson mixture model following Long et al. (2014) for estimating the effects of overall exposure effects on the marginal means. Utilizing the concept of marginalized Poisson-Poisson mixture model under cross-sectional setup (Benecha et al., 2017), we have developed a marginalized version of the Poisson-Poisson mixture model under clustered data setup following Long et al. (2015) as an extension of the Poisson-Poisson mixture model (Wang et al., 2007, 2002) for estimating the effects of overall exposure effects on the marginal means when the count responses are correlated. A generalized quasi-likelihood estimation method has also been developed under a mixture of longitudinal models for analyzing zero-inflated repeated count data. The performance of the proposed models (clustered as well as repeated measures) was investigated by conducting extensive simulation studies. Finally, real data sets have been used to illustrate the proposed models.

A comprehensive review of the cross-sectional marginalized Poisson-Poisson mixture model (Benecha et al., 2017) has been conducted with extensive simulation studies under different scenarios in Chapter 2 with a view to extending it to the longitudinal setup. Benecha et al. (2017) developed marginally-specified mean models for mixtures of two count distributions under cross-sectional setup in which the marginal parameters and the nuisance parameters may be estimated by applying the maximum likelihood (ML) technique of estimation. From the simulation studies, it was observed that the MPois-Pois mixture model offers minimal biases under the true model. The model also performs the best in terms of goodness of fit among all other competitive models in order to make inferences regarding the marginal parameters. It was also discovered that the MPois-Pois model can also be used for drawing conclusions about the overall effects of exposure on the marginal mean in situations where data arises from mixture of two susceptible classes.

The main contributions of this thesis are reported in Chapter 3 and Chapter 4 for longitudinal clustered and repeated measures data setup, respectively. In this thesis, we were able to develop a random effects marginalized Poisson-Poisson mixture (REMPois-Pois) model under the framework of ML estimation for analysing clustered count data. To address the within-cluster correlation into the model, we have considered random effects in the both latent class models. The estimation method of the proposed model was examined under two cases (i) latent class's random effects are independent and (ii) bivariate latent class's random effects. The proposed model was then fitted by employing the Gauss–Hermite quadrature method to approximate the integral while maximizing the log-likelihood function. The performance of our proposed model was examined through extensive simulation studies. The simulation studies were conducted with a different number of clusters, different cluster sizes, and a different proportion of mixture. For all of the simulation setups, the amount of bias for the marginal parameters of the fixed effects were found negligible. The random effect parameter and all the nuisance parameters had highest bias for low number of cluster and/or cluster size, and these biases were decreased with increasing the cluster size and/or with increasing the number of clusters. It was discovered that for all of the parameters, the coverage probabilities were always close to the nominal level of confidence. Similar performance of the proposed model with bivariate random effects was observed.

To illustrate the proposed REMPois-Pois model for correlated clustered count data, a nationwide representative data set extracted from the Bangladesh Demographic and Health Survey (BDHS), 2014 has been analyzed to find the potential factors associated with the number of antenatal care (ANC) visits. Note that this data set was collected using two-stage cluster sampling. In the absence of covariates, the mixing probability was estimated as 0.66. The variance components of the random effects were 0.297 (p-value < 0.001) for the marginal model and 0.431 (p-value < 0.001) for the component-1 model. From this analysis, it was found that covariates *place of residence, education level, media exposure, mother's age at birth, difference between husband and wife age (years), wealth index, birth order* have significant association with ANC visits.

To analyze zero-inflated repeated measures count data, we proposed repeated measures marginalized Poisson-Poisson mixture (RMMPois-Pois) model. As the so-called 'working' correlation based GEE technique may not provides efficient regression estimates, the regression parameters of RMMPois-Pois model have been estimated using the GQL estimation technique following Sutradhar and Das (1999). The proposed approach directly models the marginal means from mixtures of correlated counts arising from two subpopulations. This model formulation offers meaningful statements about an exposure effect on the marginal means of the count responses, in contrast to the unobserved latent class models for which the mixture model is accounted for. Extensive simulation studies were carried out to investigate the performance of our proposed RMMPois-Pois model. The simulation studies were conducted with a different number of individuals, number of occasions, proportion of mixture and longitudinal correlation coefficients. From the simulation studies, it was depicted that the estimates of regression parameters and correlation parameters for marginal model had minimal amount of biases for all the settings. It was also observed that the amount of bias of the regression parameters as well as correlation parameters for component-1 model decreases with the increase of the number of individuals. It was found that the coverage probabilities for all the parameters approximately closed to the nominal level of confidence.

To illustrate the proposed RMMPois-Pois model, a real life repeated count data set obtained from a pharmaceutical company was analyzed to examine how the treatments for a particular disease work on the number of episodes for certain side effect.

To analyze zero-inflated correlated count data, future research could extend the marginalized count regression models to mixtures of Poisson and negative binomial distributions or mixtures of two negative binomial distributions and could allow the mixing probabilities to depend on covariates. The proposed marginalized models can also be extended to deal with missing values and outliers present in correlated count data.

Appendices

Appendix A

A.1 Hessian: MPois-Pois Model

The Hessian is a $(2p+1)\times(2p+1)$ matrix of second partial derivative of log-likelihood function $\ell(\boldsymbol{\theta})$ i.e.,

$$H(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \tau^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \tau \partial \boldsymbol{\alpha}'} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \tau \partial \boldsymbol{\beta}'} \\ & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\beta}'} \\ & & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \end{bmatrix}.$$

The expression of elements of $H(\boldsymbol{\theta})$ are given below:

$$\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \tau^2} =& \sum_{i=1}^{n} \Bigg[\Bigg\{ \bigg(\Big\{ 1 - M_i^{(1)} \Big\} M_i^{(9)} + M_i^{(11)} M_i^{(1)} \bigg) M_i^{(6)} M_i^{(1)} + M_i^{(7)} \\
& - \frac{\big[M_i^{(10)}\big]^2}{M_i^{(8)}} \Bigg\} \frac{1}{M_i^{(8)}} - \frac{e^{\tau}}{\big(1+e^{\tau}\big)^2} \Bigg], \\
\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} =& \sum_{i=1}^{n} \Bigg[\boldsymbol{x}_i \Bigg\{ \bigg(\Big\{ M_i^{(5)} + y_i \Big[(y_i - 1)[M_i^{(4)}]^{y_i - 2} - 2[M_i^{(4)}]^{y_i - 1} \Big] \Big\} M_i^{(3)} + M_i^{(5)} - y_i \big[M_i^{(4)} \big]^{y_i - 1} \bigg) \\
& \times M_i^{(6)} M_i^{(3)} + \bigg(\big(y - e^{\boldsymbol{x}_i' \boldsymbol{\alpha}} \big)^2 - e^{\boldsymbol{x}_i' \boldsymbol{\alpha}} \bigg) M_i^{(7)} - \big[M_i^{(13)} \big]^2 / M_i^{(8)} \Bigg\} \frac{1}{M_i^{(8)}} \boldsymbol{x}_i' \Bigg], \\
\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} =& \sum_{i=1}^{n} \Bigg[\boldsymbol{x}_i \frac{\bigg(\Big\{ 1 - M_i^{(2)} \Big\} M_i^{(9)} + M_i^{(2)} \Big\{ M_i^{(11)} - M_i^{(6)} \big[M_i^{(9)} \big]^2 / M_i^{(8)} \Big\} \bigg) \big(1 + e^{\tau}\big) M_i^{(6)} e^{\boldsymbol{x}_i' \boldsymbol{\beta}}}{M_i^{(8)}} \boldsymbol{x}_i' \Bigg]
\end{aligned}$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\tau\partial\boldsymbol{\alpha}'} = \sum_{i=1}^{n}\left[\frac{\left(M_i^{(5)} + M_i^{(1)}M_i^{(12)} - y_i\left[M_i^{(4)}\right]^{y_i-1}\right)M_i^{(6)}M_i^{(3)} + M_i^{(7)}\left(y_i - e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}\right) - M_i^{(14)}}{M_i^{(8)}}\right]\boldsymbol{x}_i',$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\tau\partial\boldsymbol{\beta}'} = \sum_{i=1}^{n}\left[\left\{\left(\left\{1 - (1+\frac{1}{e^{\tau}})M_i^{(1)}\right\}M_i^{(9)} + M_i^{(11)}(1+\frac{1}{e^{\tau}})M_i^{(1)}\right)e^{\tau} - \frac{(1+e^{\tau})M_i^{(10)}M_i^{(9)}}{M_i^{(8)}}\right\}\right.$$
$$\left.\times M_i^{(6)}e^{\boldsymbol{x}_i'\boldsymbol{\beta}}/M_i^{(1)}\right]\boldsymbol{x}_i',$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}'} = \sum_{i=1}^{n}\left[\boldsymbol{x}_i\frac{(1+e^{\tau})M_i^{(6)}\left(M_i^{(3)}M_i^{(12)} - M_i^{(13)}M_i^{(9)}/M_i^{(8)}\right)e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{M_i^{(8)}}\boldsymbol{x}_i'\right],$$

where

$$M_i^{(1)} = e^{\tau}\left(e^{\boldsymbol{x}_i'\boldsymbol{\beta}} - e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}\right),$$

$$M_i^{(2)} = (1+e^{\tau})e^{\boldsymbol{x}_i'\boldsymbol{\beta}},$$

$$M_i^{(3)} = e^{\tau}e^{\boldsymbol{x}_i'\boldsymbol{\alpha}},$$

$$M_i^{(4)} = M_i^{(2)} - M_i^{(3)},$$

$$M_i^{(5)} = \left[M_i^{(4)}\right]^{y_i},$$

$$M_i^{(6)} = e^{-M_i^{(4)}},$$

$$M_i^{(7)} = e^{\boldsymbol{x}_i'\boldsymbol{\alpha}y_i + \tau - e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}},$$

$$M_i^{(8)} = M_i^{(5)}M_i^{(6)} + M_i^{(7)},$$

$$M_i^{(9)} = y_i\left[M_i^{(4)}\right]^{y_i-1} - M_i^{(5)},$$

$$M_i^{(10)} = M_i^{(6)}M_i^{(1)}M_i^{(9)} + M_i^{(7)},$$

$$M_i^{(11)} = y_i\left\{(y_i-1)\left[M_i^{(4)}\right]^{y_i-2} - \left[M_i^{(4)}\right]^{y_i-1}\right\},$$

$$M_i^{(12)} = y_i\left\{2\left[M_i^{(4)}\right]^{y_i-1} - (y_i-1)\left[M_i^{(4)}\right]^{y_i-2}\right\} - M_i^{(5)},$$

$$M_i^{(13)} = \left(M_i^{(5)} - y_i\left[M_i^{(4)}\right]^{y_i-1}\right)M_i^{(6)}M_i^{(3)} + M_i^{(7)}\left(y_i - e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}\right),$$

$$M_i^{(14)} = M_i^{(13)}M_i^{(10)}/M_i^{(8)}.$$

A.2 The EM Algorithm: Poisson-Poisson Mixture Model

The parameters of Poisson-Poisson mixture model can be estimated by the EM algorithm (Benecha et al., 2017; McLachlan and Peel, 2000). In practice, the value of D_i for Eq. (2.14) is not known. Suppose that (Y_i, D_i) is the complete data vector for ith observation $(i = 1, 2, \cdots, n)$ and $f_g(\cdot), g = 1, 2$ is the probability distribution of the count response for gth component. Under Poisson-Poisson mixture distribution, each y_i can be thought as having arisen from one of the components of mixture where $f_1(y_i; \cdot) \equiv \text{Pois}(\mu_{1,i})$, $f_2(y_i; \cdot) \equiv \text{Pois}(\mu_{2,i})$ as shown in Eq.(2.14) for which we are intended to find estimates of associated parameters as given in Eq.(2.17). If we assume same set of covariates in the both models of mixing component, then under the assumption of independent observations, the log-likelihood function of parameter, $\boldsymbol{\theta} = (\pi^*, \boldsymbol{\alpha}', \boldsymbol{\gamma}')'$ for complete data can be written as

$$\begin{aligned}\ell_c(\pi^*, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = &\sum_{i=1}^{n} \Big[d_i \text{logit}(\pi^*) + \log(1-\pi^*)\Big] \\ &+ \sum_{i=1}^{n} d_i \Big[y_i \boldsymbol{x_i'\alpha} - \exp(\boldsymbol{x_i'\alpha}) - \log(y_i!)\Big] \\ &+ \sum_{i=1}^{n} (1-d_i) \Big[y_i \boldsymbol{x_i'\gamma} - \exp(\boldsymbol{x_i'\gamma}) - \log(y_i!)\Big].\end{aligned} \tag{A.1}$$

The EM algorithm is applied to this problem for finding the MLE of $(\pi^*, \boldsymbol{\alpha}', \boldsymbol{\gamma}')'$ by treating d_i as missing. It proceeds iteratively in two steps, E (for expectation) and M (for maximization).

E-Step:

In EM framework, the latent data are handled in the E-step. It computes the conditional expectation of the complete data log likelihood, $\ell_c(\pi^*, \boldsymbol{\alpha}, \boldsymbol{\gamma})$, given the observed data $\boldsymbol{y}$, using the current fit $\boldsymbol{\theta}^{(0)}$, where $\boldsymbol{\theta} = (\pi^*, \boldsymbol{\alpha}', \boldsymbol{\gamma}')'$. Therefore, given the initial values

$\boldsymbol{\theta}^{(0)} = (\pi^{*(0)}, \boldsymbol{\alpha}'^{(0)}, \boldsymbol{\gamma}'^{(0)})'$, the E-step computes

$$E\Big[\ell_c(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)})\Big] = \sum_{i=1}^{n} \Big(E\big[D_i|\boldsymbol{\theta}^{(0)}, y_i, \boldsymbol{x_i}\big]\text{logit}(\pi^*) + \log(1-\pi^*)\Big)$$
$$+ \sum_{i=1}^{n} E\Big[D_i|\boldsymbol{\theta}^{(0)}, y_i, \boldsymbol{x_i}\Big]\Big[y_i\boldsymbol{x_i'\alpha} - \exp(\boldsymbol{x_i'\alpha}) - \log(y_i!)\Big] \tag{A.2}$$
$$+ \sum_{i=1}^{n} \Big[1 - E[D_i|\boldsymbol{\theta}^{(0)}, y_i, \boldsymbol{x_i}]\Big]\Big[y_i\boldsymbol{x_i'\gamma} - \exp(\boldsymbol{x_i'\gamma}) - \log(y_i!)\Big].$$

As the complete data log-likelihood in Eq.(A.1) is linear in the unobservable data d_i, the E-step simply requires the calculation of the current conditional expectation of D_i, given the observed data y_i. Therefore we can write

$$E_{\boldsymbol{\theta}^{(0)}}\Big[D_i|y_i, \boldsymbol{\theta}^{(0)}\Big] = \frac{\pi^{*(0)} f_1(y_i; \boldsymbol{\alpha}')}{\pi^{*(0)} f_1(y_i; \boldsymbol{\alpha}') + (1-\pi^{*(0)}) f_2(y_i; \boldsymbol{\gamma}')} \equiv P_i^{(0)}, \tag{A.3}$$

for $i = 1, 2, \cdots, n$. Using Eq.(A.3), the conditional expectation in Eq.(A.2) can be written as

$$E\Big[\ell_c(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)})\Big] = \Bigg[\sum_{i=1}^{n} P_i^{(0)}\text{logit}(\pi) + n\log(1-\pi)\Bigg]$$
$$+ \Bigg[\sum_{i=1}^{n} P_i^{(0)}\Big(y_i\boldsymbol{x_i}'\alpha - \exp(\boldsymbol{x_i}'\alpha) - \log(y_i!)\Big)\Bigg]$$
$$+ \Bigg[\sum_{i=1}^{n} \Big(1 - P_i^{(0)}\Big)\Big(y_i\boldsymbol{x_i}'\gamma - \exp(\boldsymbol{x_i}'\gamma) - \log(y_i!)\Big)\Bigg]$$
$$= \ell_{\pi^*} + \ell_{\boldsymbol{\alpha}} + \ell_{\boldsymbol{\gamma}}. \tag{A.4}$$

M-Step:

The M-step maximizes Eq.(A.4). To obtain the next estimates in the M step, the three components $\ell_{\pi^*}, \ell_{\boldsymbol{\alpha}}$ and $\ell_{\boldsymbol{\gamma}}$ in Eq.(A.4) can be optimized separately. Maximizing ℓ_{π^*} with respect to π^* of Eq.(A.4) gives the next updated estimate of π^* as

$$\pi^{*(1)} = \sum_{i=1}^{n} \frac{P_i^{(0)}}{n} \tag{A.5}$$

It is clear that both components $\ell_{\boldsymbol{\alpha}}$ and $\ell_{\boldsymbol{\gamma}}$ of Eq.(A.4) correspond to the weighted log-likelihood of standard Poisson regression model, a member of generalized linear models (GLMs). To estimate the parameters of both components, one can perform Fisher's method of scoring separately. For a GLM, it is equivalent to use iteratively reweighed least squares (IRLS) method. Thus, one can apply IRLS method separately to get estimate of $\boldsymbol{\alpha}^{(1)}$ and $\boldsymbol{\gamma}^{(1)}$ in the first step. Utilizing the parameters $\pi^{*(1)}$, $\boldsymbol{\alpha}^{(1)}$ and $\boldsymbol{\gamma}^{(1)}$ in the first step, EM again computes and optimizes the expected log-likelihood and continues iterations between the two steps until convergence.

It follows that on the $(r+1)$th iteration, E-step requires the calculation of $E\Big[\ell_c(\pi^*, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\boldsymbol{\theta}^{(r)})\Big]$, where $\boldsymbol{\theta}^{(r)}$ is the value of $\boldsymbol{\theta}$ after the rth EM iteration. Then the $(r+1)$th iteration requires the global maximization of Eq.(A.4) with respect to $\boldsymbol{\theta}$ over the parameter space $\boldsymbol{\Omega}$ to give the updated estimate of $\boldsymbol{\theta}^{(r+1)}$.

Appendix B

B.1 Gauss-Hermite numerical quadrature for evaluating integrals

Gaussian quadrature is used to approximate integrals of functions with respect to a given kernel by a weighted average of the integrand evaluated at predetermined abscissas. The weights and abscissas used in Gaussian quadrature rules for the most common kernels can be obtained from the tables of Abramowitz et al. (1964) or by using an algorithm proposed by Golub (1973). To approximate a two-dimensional integrals with the Gaussian kernel, the Gauss-Hermite quadrature (GHQ) can be expressed as follows

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u,c)e^{-(u^2+c^2)}dudc \approx \sum_{k=1}^{m}\sum_{l=1}^{m} w_k^u w_l^c f(u_k,c_l), \tag{B.1}$$

where u_k and c_l are abscissas of m-point GHQ corresponding to u and c, respectively with associated respective weights w_k^u, w_l^c, $k = l = 1,\ldots,m$. The log-likelihood function of REMPois-Pois model as given in Eq.(3.7) can be written as

$$\ell(\boldsymbol{\delta}) = \sum_{i=1}^{K}\log\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{j=1}^{n_i} f(y_{ij}|u_i,c_i;\boldsymbol{\delta})\frac{1}{\sigma_u\sigma_c 2\pi\sqrt{1-\rho^{*2}}}e^{-\frac{1}{2(1-\rho^{*2})}\left((\frac{u_i}{\sigma_u})^2+(\frac{c_i}{\sigma_c})^2-2\rho^*(\frac{u_i}{\sigma_u})(\frac{c_i}{\sigma_c})\right)}du_idc_i\right]. \tag{B.2}$$

Change-of-variable operations have been carried out at this stage. For this purpose, let $u_i = \sqrt{2(1-\rho^{*2})}\sigma_u u_i^*$ and $c_i = \sqrt{2(1-\rho^{*2})}\sigma_c c_i^*$. Then the log-likelihood function becomes

$$\begin{aligned}\ell(\boldsymbol{\delta}) =& \sum_{i=1}^{K}\left[\log\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{j=1}^{n_i} f\Big(y_{ij}|\sqrt{2(1-\rho^{*2})}\sigma_u u_i^*, \sqrt{2(1-\rho^{*2})}\sigma_c c_i^*; \boldsymbol{\delta}\Big)\right.\right.\\ &\left.\left.\times\Big(\frac{\sqrt{1-\rho^{*2}}}{\pi}\Big)e^{-\left(u_i^{*2}+c_i^{*2}-2\rho^* u_i^* c_i^*\right)}du_i^* dc_i^*\right\}\right]\\ =& \sum_{i=1}^{K}\left[\log\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big(\frac{\sqrt{1-\rho^{*2}}e^{2\rho^* u_i^* c_i^*}}{\pi}\Big)\right.\right.\\ &\left.\left.\times\prod_{j=1}^{n_i} f\Big(y_{ij}|\sqrt{2(1-\rho^{*2})}\sigma_u u_i^*, \sqrt{2(1-\rho^{*2})}\sigma_c c_i^*; \boldsymbol{\delta}\Big)e^{-\left(u_i^{*2}+c_i^{*2}\right)}\right\}\right]. \qquad (B.3)\end{aligned}$$

Using Eq.(B.1), the GHQ rule for two-dimensional integrals given in Eq.(B.3) can be approximated by

$$\begin{aligned}\ell(\boldsymbol{\delta}|\boldsymbol{y}_i) \approx& \sum_{i=1}^{K}\left[\log\left\{\sum_{l_1=1}^{m}\sum_{l_2=1}^{m}\frac{w_{l_1}^u w_{l_2}^c\sqrt{1-\rho^{*2}}}{\pi}e^{2\rho^* q_{l_1}^u q_{l_2}^c}\right.\right.\\ &\left.\left.\times\prod_{j=1}^{n_i} f\Big(y_{ij}|\sqrt{2(1-\rho^{*2})}\sigma_u q_{l_1}^u, \sqrt{2(1-\rho^{*2})}\sigma_c q_{l_2}^c; \boldsymbol{\delta}\Big)\right\}\right], \qquad (B.4)\end{aligned}$$

where $q_{l_1}^u$ and $q_{l_2}^c$are abscissas of m-point GHQ corresponding to u^* and c^*, respectively with associated respective weights $w_{l_1}^u$, $w_{l_2}^c$, $l_1 = l_2 = 1, \cdots, m$.

B.2 REMPois-Pois Model: Score function

The elements of score function for the REMPois-Pois model can be computed as

$$U_1 = \sum_{i=1}^{K}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m}\frac{w^u_{l_1}w^c_{l_2}}{D_i}\frac{1}{\Big(\prod_{m=1}^{n_i} y_{im}!\Big)\pi\big(1+e^{\tau}\big)^{n_i}}\Bigg[\sum_{j=1}^{n_i}\Bigg\{\Bigg(e^{\tau+\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}} - e^{\tau+\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sqrt{2}\sigma_c q^c_{l_2}}\Bigg)$$

$$\Bigg(\bigg[\big(1+e^{\tau}\big)e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}} - e^{\tau+\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\bigg]^{y_{ij}} - y_{ij}\bigg[\big(1+e^{\tau}\big)e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}} - e^{\tau+\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\bigg]^{y_{ij}-1}\Bigg)$$

$$\Bigg(e^{\Big[e^{\tau+\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}} - (1+e^{\tau})e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}}\Big]}\Bigg) + e^{\Big[\tau+y_{ij}(\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}) - e^{\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\Big]}\Bigg\}$$

$$\Bigg\{\prod_{k\neq j=1}^{n_i}\Bigg(\bigg[\big(1+e^{\tau}\big)e^{\boldsymbol{x}'_{ik}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}} - e^{\tau+\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\bigg]^{y_{ik}} e^{\Big[e^{\tau+\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}} - (1+e^{\tau})e^{\boldsymbol{x}'_{ik}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}}\Big]} +$$

$$e^{\Big[\tau+y_{ik}(\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}) - e^{\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\Big]}\Bigg)\Bigg\}$$

$$-\frac{n_i}{\big(1+e^{\tau}\big)}\prod_{k=1}^{n_i}\Bigg(\bigg[\big(1+e^{\tau}\big)e^{\boldsymbol{x}'_{ik}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}} - e^{\tau+\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\bigg]^{y_{ik}} e^{\Big[e^{\tau+\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}} - (1+e^{\tau})e^{\boldsymbol{x}'_{ik}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}}\Big]}$$

$$+\, e^{\Big[\tau+y_{ik}(\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c c^{*}_{l_2}) - e^{\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\Big]}\Bigg)\Bigg],$$

$$U_2 = \sum_{i=1}^{K}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m}\frac{w^u_{l_1}w^c_{l_2}}{D_i}\frac{1}{\Big(\prod_{m=1}^{n_i} y_{im}!\Big)\pi\big(1+e^{\tau}\big)^{n_i}}\Bigg[\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}\Bigg\{e^{\Big[e^{\tau+\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}} - (1+e^{\tau})e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}}\Big]}$$

$$\Bigg(\bigg[\big(1+e^{\tau}\big)e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}} - e^{\tau+\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\bigg]^{y_{ij}} - y_{ij}\bigg[\big(1+e^{\tau}\big)e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}} - e^{\tau+\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\bigg]^{y_{ij}-1}\Bigg)$$

$$e^{\tau+\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}} + e^{\Big[\tau+y_{ij}(\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}) - e^{\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\Big]}\Bigg(y_{ij} - e^{\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\Bigg)\Bigg\}$$

$$\Bigg\{\prod_{k\neq j=1}^{n_i}\Bigg(e^{\Big[\tau+y_{ik}(\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}) - e^{\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\Big]} +$$

$$\bigg[\big(1+e^{\tau}\big)e^{\boldsymbol{x}'_{ik}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}} - e^{\tau+\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}}\bigg]^{y_{ik}} e^{\Big[e^{\tau+\boldsymbol{x}'_{ik}\boldsymbol{\alpha}+\sqrt{2}\sigma_c q^c_{l_2}} - (1+e^{\tau})e^{\boldsymbol{x}'_{ik}\boldsymbol{\beta}+\sqrt{2}\sigma_u q^u_{l_1}}\Big]}\Bigg)\Bigg\}\Bigg],$$

$$U_3 = \sum_{i=1}^{K}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m}\frac{w_{l_1}^u w_{l_2}^c c_{l_2}^*}{D_i}\frac{\sqrt{2}}{\left(\prod_{m=1}^{n_i} y_{im}!\right)\pi\left(1+e^{\tau}\right)^{n_i}}\left[\sum_{j=1}^{n_i}\left\{e^{\left[e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}-(1+e^{\tau})e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}\right]}\right.\right.$$

$$\left(\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}-e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]^{y_{ij}}-y_{ij}\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}-e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]^{y_{ij}-1}\right)$$

$$\left.e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}+\left(y_{ij}-e^{\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right)\right\}$$

$$\left\{\prod_{k\neq j=1}^{n_i}\left(\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ik}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}-e^{\tau+\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]^{y_{ik}}e^{\left[e^{\tau+\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}-(1+e^{\tau})e^{\boldsymbol{x}_{ik}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}\right]}+\right.\right.$$

$$\left.\left.\left.e^{\left[\tau+y_{ik}(\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c)-e^{\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]}\right)\right\}\right],$$

$$U_4 = \sum_{i=1}^{K}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m}\frac{w_{l_1}^u w_{l_2}^c}{D_i}\frac{1}{\left(\prod_{m=1}^{n_i} y_{im}!\right)\pi\left(1+e^{\tau}\right)^{n_i-1}}\left[\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}\left\{e^{\left[e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}-(1+e^{\tau})e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}\right]}\right.\right.$$

$$\left(y_{ij}\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}-e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]^{y_{ij}-1}-\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}-e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]^{y_{ij}}\right)$$

$$\left.e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}\right\}\left\{\prod_{k\neq j=1}^{n_i}\left(e^{\left[\tau+y_{ik}(\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c)-e^{\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]}+\right.\right.$$

$$\left.\left.\left.\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ik}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}-e^{\tau+\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]^{y_{ik}}e^{\left[e^{\tau+\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}-(1+e^{\tau})e^{\boldsymbol{x}_{ik}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}\right]}\right)\right\}\right],$$

$$U_5 = \sum_{i=1}^{K}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m}\frac{w_{l_1}^u u_{l_1}^* w_{l_2}^c}{D_i}\frac{\sqrt{2}}{\left(\prod_{m=1}^{n_i} y_{im}!\right)\pi\left(1+e^{\tau}\right)^{n_i-1}}\left[\sum_{j=1}^{n_i}\left\{e^{\left[e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}-(1+e^{\tau})e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}\right]}\right.\right.$$

$$\left(y_{ij}\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}-e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]^{y_{ij}-1}-\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}-e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]^{y_{ij}}\right)$$

$$\left.e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}\right\}\left\{\prod_{k\neq j=1}^{n_i}\left(e^{\left[\tau+y_{ik}(\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c)-e^{\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]}+\right.\right.$$

$$\left.\left.\left.\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ik}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}-e^{\tau+\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]^{y_{ik}}e^{\left[e^{\tau+\boldsymbol{x}_{ik}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}-(1+e^{\tau})e^{\boldsymbol{x}_{ik}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}\right]}\right)\right\}\right],$$

where

$$D_i = \sum_{l_1=1}^{m}\sum_{l_2=1}^{m}\frac{w_{l_1}^u w_{l_2}^c}{\pi}\prod_{j=1}^{n_i}\Bigg\{\frac{1}{y_{ij}!\big(1+e^{\tau}\big)}\Bigg(e^{\left[\tau+y_{ij}(\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c)-e^{\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]}$$
$$+\,e^{-\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}-e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]}\left[\left(1+e^{\tau}\right)e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\sqrt{2}\sigma_u q_{l_1}^u}-e^{\tau+\boldsymbol{x}_{ij}'\boldsymbol{\alpha}+\sqrt{2}\sigma_c q_{l_2}^c}\right]^{y_{ij}}\Bigg)\Bigg\}.$$

B.3 The EM Algorithm: Random Effects Poisson-Poisson Mixture Model

Let $Y_{ij}(i=1,\cdots,K; j=1,\cdots,n_i)$ be the count variable of interest for the jth individual in the ith cluster. Since the ith cluster-population is divided into component-1 and component-2 and for a given random effect, the observations of component-1 follow independent Poisson distributions and that of component-2 follow independent another Poisson distributions. Suppose that $\mu_{1,ij}^* = E\big[Y_{ij}|c_i\big]$ (c_i is a random effect for component-1) and $\mu_{2,ij}^* = E\big[Y_{ij}|d_i\big]$ (d_i is a random effect for component-2). Let $b_i^T = (c_i, d_i) \sim N_2(\mathbf{0}, \Sigma^*)$, where Σ^* is a 2×2 covariance matrix with diagonal elements σ_c^2, σ_d^2 and off-diagonal element $\rho^{**}\sigma_c\sigma_d$.

The EM algorithm for random effects Poisson-Poisson mixture model is derived following Hall (2000). Let $\boldsymbol{\delta^*} = (\boldsymbol{\alpha}', \sigma_c, \boldsymbol{\gamma}', \sigma_d, \pi^*, \rho^{**})'$ be the combined vector of model parameters. Under the regression setup as in Eq.(3.1), the likelihood function of the Poisson-Poisson mixture model with random effects is given as

$$\mathrm{L}(\boldsymbol{\delta}^*|\boldsymbol{y}) = \prod_{i=1}^{K}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{j=1}^{n_i} f(y_{ij}|b_i;\boldsymbol{\alpha},\boldsymbol{\gamma})f(b_i|\sigma_c,\sigma_d,\rho^{**})dc_i dd_i, \tag{B.5}$$

where $f(y_{ij}|c_i, d_i; \boldsymbol{\alpha}, \boldsymbol{\gamma})$ is the pmf of Poisson-Poisson mixture distribution for analyzing clustered data defined as

$$f(Y_{ij}=y_{ij}|\boldsymbol{\alpha}, c_i, \boldsymbol{\gamma}, d_i, \tau) = \pi^*\frac{e^{-\mu_{1,ij}^*}{\mu_{1,ij}^*}^{y_{ij}}}{y_{ij}!} + (1-\pi^*)\frac{e^{-\mu_{2,ij}^*}{\mu_{2,ij}^*}^{y_{ij}}}{y_{ij}!}, \tag{B.6}$$

and $f(b_i|\sigma_c, \sigma_d, \rho^{**})$ is a pdf of bivariate normal distribution. For avoiding complexity arisen in computation as well as notation, we assume that $b_i^* = (c_i^*, d_i^*)'$ with $c_i^* = c_i/\sigma_c, d_i^* = d_i/\sigma_d$; and that c_i and d_i are independent i.e., $\rho^{**} = 0$. Then Eq.(B.5) can be written as

$$\mathrm{L}(\boldsymbol{\delta}^*|\boldsymbol{y}) = \prod_{i=1}^{K} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f(y_{ij}|b_i^*; \boldsymbol{\delta}^*) \phi(c_i^*) \phi(d_i^*) dc_i^* dd_i^*, \tag{B.7}$$

where

$$f(y_{ij}|b_i; \boldsymbol{\delta}^*) = \pi^* \frac{e^{-\mu_{1,ij}^*} {\mu_{1,ij}^*}^{y_{ij}}}{y_{ij}!} + (1 - \pi^*) \frac{e^{-\mu_{2,ij}^*} {\mu_{2,ij}^*}^{y_{ij}}}{y_{ij}!}, \tag{B.8}$$

and $\phi(\cdot)$ is the standard normal pdf, and

$$\log(\mu_{1,ij}^*) = \boldsymbol{x}_{ij}' \boldsymbol{\alpha} + \sigma_c c_i^*; \quad \log(\mu_{2,ij}^*) = \boldsymbol{x}_{ij}' \boldsymbol{\gamma} + \sigma_d d_i^*, \tag{B.9}$$

where $\boldsymbol{x}_{ij}$ is the design matrix for the jth individual of the ith cluster in both the components of the mixture. The method of estimating regression parameters and random effects parameters by maximizing Eq.(B.7) is usually complicated. The complication can be reduced to a great extent by employing the EM algorithm. In the EM framework for Poisson-Poisson mixture model with random effects, the observed count within the formulation of the mixture problem and also the random effects are viewed as being incomplete. In other words, each y_{ij} can be thought as having arisen from one of the components of the mixture distribution Eq.(B.8) for which we are intended to find estimates of associated parameters as given in equation Eq.(B.9). But the EM problem at this stage is that in Eq.(B.9) the random effects c_i or d_i, associated with the component of mixture, can also be thought as missing.

Let us consider the latent Bernoulli variable $u_{ij}, i = 1, \cdots, K, j = 1, \cdots, n_i$ denote the component membership. i.e., $u_{ij} = 1$ if Y_{ij} is drawn following the distribution of component-1 and $u_{ij} = 0$ if it is from component-2. The complete data are then $(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{b})$ for the EM algorithm, where $(\boldsymbol{u}, \boldsymbol{b})$ are thought as missing data. Let $f_1(y_{ij}) \equiv \mathrm{Pois}(\mu_{1,ij}^*)$,

$f_2(y_{ij}) \equiv \text{Pois}(\mu^*_{2,ij})$ and $b^*_i \sim N_2(\mathbf{0}, I)$, where I is a identity matrix and that b^*_i has pdf $\phi_2(b^*_i)$. The likelihood of Poisson-Poisson mixture model with random effects based on the complete data $(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{b})$ under the assumption of independence of $\boldsymbol{y}_i, i = 1, \cdots, K$ and also independence of $b_i, i = 1, \cdots, K$ can be written as

$$
\begin{aligned}
L_c(\boldsymbol{\delta}^*|\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{b}) &= \prod_{i=1}^{K} Pr(\boldsymbol{Y}_i = \boldsymbol{y}_i, \boldsymbol{U}_i = \boldsymbol{u}_i, B_i = b_i|\boldsymbol{\delta}^*; \boldsymbol{x}_{ij}) \\
&= \prod_{i=1}^{K} Pr(\boldsymbol{Y}_i = \boldsymbol{y}_i, \boldsymbol{U}_i = \boldsymbol{u}_i|B_i = b_i, \boldsymbol{\delta}^*; \boldsymbol{x}_{ij})\phi_2(b^*_i) \\
&= \left[\prod_{i=1}^{K}\left\{\prod_{j=1}^{n_i} Pr(Y_{ij} = y_{ij}, U_{ij} = u_{ij}|B_i = b_i, \boldsymbol{\delta}^*; \boldsymbol{x}_{ij})\right\}\phi_2(b^*_i)\right] \\
&= \left[\prod_{i=1}^{K}\phi_2(b^*_i)\right]\left[\prod_{i=1}^{K}\prod_{j=1}^{n_i} Pr(Y_{ij} = y_{ij}|U_{ij} = u_{ij}, B_i = b_i, \boldsymbol{\delta}^*; \boldsymbol{x}_{ij})\right. \\
&\quad \left. \times Pr(U_{ij} = u_{ij}|B_i = b_i, \pi^*)\right] \\
&= \left[\prod_{i=1}^{K}\phi_2(b^*_i)\right]\left[\prod_{i=1}^{K}\prod_{j=1}^{n_i} Pr(Y_{ij} = y_{ij}|U_{ij} = u_{ij}, B_i = b_i, \boldsymbol{\delta}^*; \boldsymbol{x}_{ij})Pr(U_{ij} = u_{ij}|\pi^*)\right] \\
&= \left[\prod_{i=1}^{K}\phi_2(b^*_i)\right] \\
&\quad \times \left[\prod_{i=1}^{K}\prod_{j=1}^{n_i}\left[f_1(y_{ij}|\boldsymbol{\alpha}, \boldsymbol{x}_{ij}, c_i, \sigma_c)\right]^{u_{ij}}\left[f_2(y_{ij}|\boldsymbol{\gamma}, \boldsymbol{x}_{ij}, d_i, \sigma_d)\right]^{1-u_{ij}}\left[\pi^*\right]^{u_{ij}}\left[1-\pi^*\right]^{1-u_{ij}}\right] \\
&= \left[\prod_{i=1}^{K}\phi_2(b^*_i)\right]\left[\prod_{i=1}^{K}\prod_{j=1}^{n_i}\left\{\pi^* f_1(y_{ij}|\boldsymbol{\alpha}, \boldsymbol{x}_{ij}, c_i, \sigma_c)\right\}^{u_{ij}}\left\{(1-\pi^*)f_2(y_{ij}|\boldsymbol{\gamma}, \boldsymbol{x}_{ij}, d_i, \sigma_d)\right\}^{1-u_{ij}}\right] \\
&= \left[\prod_{i=1}^{K}\phi_2(b^*_i)\right]\left[\prod_{i=1}^{K}\prod_{j=1}^{n_i}\left\{\pi^*\frac{e^{-\mu^*_{1,ij}}{\mu^*_{1,ij}}^{y_{ij}}}{y_{ij}!}\right\}^{u_{ij}}\left\{(1-\pi^*)\frac{e^{-\mu^*_{2,ij}}{\mu^*_{2,ij}}^{y_{ij}}}{y_{ij}!}\right\}^{1-u_{ij}}\right].
\end{aligned}
\tag{B.10}
$$

The corresponding log-likelihood is given as

$$\begin{aligned}
\ell_c(\boldsymbol{\delta}^*|\boldsymbol{y},\boldsymbol{u},\boldsymbol{b}^*) =& \log f(\boldsymbol{b}^*) + \log f(\boldsymbol{u}|\boldsymbol{b}^*;\boldsymbol{\delta}^*) + \log f(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{b}^*;\boldsymbol{\delta}^*) \\
=& \sum_{i=1}^{K}\Big[\Big\{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}c_i^{*2}}\Big\}\Big\{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_i^{*2}}\Big\}\Big]+ \\
& \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[u_{ij}\log(\pi^*) + (1-u_{ij})\log(1-\pi^*)\Big]+ \\
& \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[u_{ij}\Big\{\log\big(\frac{1}{y_{ij}!}\big) + y_{ij}\big(\boldsymbol{x}'_{ij}\boldsymbol{\alpha} + \sigma_c c_i^*\big) - e^{\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sigma_c c_i^*}\Big\}\Big]+ \\
& \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\Big(1-u_{ij}\Big)\Big\{\log\big(\frac{1}{y_{ij}!}\big) + y_{ij}\big(\boldsymbol{x}'_{ij}\boldsymbol{\gamma} + \sigma_d d_i^*\big) - e^{\boldsymbol{x}'_{ij}\boldsymbol{\gamma}+\sigma_d d_i^*}\Big\}\Big], \quad \text{(B.11)}
\end{aligned}$$

The EM algorithm is applied to this problem for finding the MLE of $\boldsymbol{\delta}^*$ by treating c_i^*, d_i^* and u_{ij} as missing. Given a starting value for the vector of parameters $\boldsymbol{\delta}^*$, say $\boldsymbol{\delta}^{*(0)}$, the EM algorithm proceeds iteratively to obtain the required MLE by switching between an E-step (for expectation) and a M-step (for maximization)

E-Step:

In EM framework, the latent data is handled by the E-step. It follows that on the $(r+1)$th iteration, E-step requires the calculation of

$$Q(\boldsymbol{\delta}^*|\boldsymbol{\delta}^{*(r)}) = E\Big[\log f\big(\boldsymbol{y},\boldsymbol{u},\boldsymbol{b}^*|\boldsymbol{\delta}^{*(r)}\big)|\boldsymbol{y},\boldsymbol{\delta}^{*(r)}\Big] = E\Big[E\Big\{\log f\big(\boldsymbol{y},\boldsymbol{u},\boldsymbol{b}^*|\boldsymbol{\delta}^{*(r)}\big)|\boldsymbol{y},\boldsymbol{b}^*,\boldsymbol{\delta}^{*(r)}\Big\}|\boldsymbol{y},\boldsymbol{\delta}^{*(r)}\Big],$$

where the inner expectation is with respect to $\boldsymbol{u}$ only. Since $\log f(\boldsymbol{y},\boldsymbol{u},\boldsymbol{b}^*|\boldsymbol{\delta}^*)$ is linear with respect to $\boldsymbol{u}$, this inner expectation can be taken simply by substituting $\boldsymbol{u}^{(r)} = E\Big[\boldsymbol{u}|\boldsymbol{y},\boldsymbol{b}^*,\boldsymbol{\delta}^{*(r)}\Big]$ for $\boldsymbol{u}$. In this case, it is noted that $\boldsymbol{u}^{(r)}$ depends only on $\boldsymbol{b}^*$ and thus it can be expressed as a function of $\boldsymbol{b}^*$. Then the vector $\boldsymbol{u}^{(r)}(\boldsymbol{b}^*)$ can be easily computed with elements

$$\begin{aligned}
u_{ij}^{(r)}(b_i^*) &= E\Big[u_{ij}|y_{ij}, b_i^*, \boldsymbol{\delta}^{*(r)}\Big] \\
&= 1 \times Pr\Big[u_{ij} = 1|y_{ij}, b_i^*, \boldsymbol{\delta}^{*(r)}\Big] + 0 \times Pr\Big[u_{ij} = 0|y_{ij}, b_i^*, \boldsymbol{\delta}^{*(r)}\Big] \\
&= Pr\Big[u_{ij} = 1|y_{ij}, b_i^*, \boldsymbol{\delta}^{*(r)}\Big] \\
&= \frac{Pr(u_{ij} = 1, y_{ij}|b_i^*, \boldsymbol{\delta}^{*(r)})}{f(y_{ij}|b_i^*, \boldsymbol{\delta}^{*(r)})} \\
&= \frac{Pr(u_{ij} = 1|\boldsymbol{\delta}^{*(r)}) \times f(y_{ij}|u_{ij} = 1, b_i^*, \boldsymbol{\delta}^{*(r)})}{f(y_{ij}|b_i^*, \boldsymbol{\delta}^{*(r)})} \\
&= \frac{\pi^{*(r)} f_1(y_{ij}|c_i^*, \boldsymbol{\delta}^{*(r)})}{\pi^{*(0)} f_1(y_{ij}|c_i^*, \boldsymbol{\delta}^{*(r)}) + (1 - \pi^{*(0)}) f_2(y_{ij}|d_i^*, \boldsymbol{\delta}^{*(r)})} \\
&= \left[1 + \frac{(1 - \pi^{*(0)})}{\pi^{*(0)}} \frac{f_2(y_{ij}|d_i^*, \boldsymbol{\delta}^{*(r)})}{f_1(y_{ij}|c_i^*, \boldsymbol{\delta}^{*(r)})}\right]^{-1}.
\end{aligned} \tag{B.12}$$

To complete the E-step, it is required to take the outer expectation and it also requires dropping the terms not involving $\boldsymbol{\delta}^*$ and thus we obtain

$$\begin{aligned}
Q(\boldsymbol{\delta}^*|\boldsymbol{\delta}^{*(r)}) =& E\Big[\log f\big(\boldsymbol{y}, \boldsymbol{u}^{(r)}(\boldsymbol{b}^*), \boldsymbol{b}^*|\boldsymbol{\delta}^*\big)|\boldsymbol{y}, \boldsymbol{\delta}^{*(r)}\Big] \\
=& \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \ell_c\big(\boldsymbol{\delta}^*|y_{ij}, u_{ij}^{(r)}(b_i^*)\big) f\big(\boldsymbol{y}_i|b_i^*, \boldsymbol{\delta}^{*(r)}\big)\phi_2(b_i^*) dc_i^* dd_i^*}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f\big(\boldsymbol{y}_i|b_i^*, \boldsymbol{\delta}^{*(r)}\big)\phi_2(b_i^*) dc_i^* dd_i^*}.
\end{aligned} \tag{B.13}$$

Here, we have used the fact that

$$f\big(b_i^*|\boldsymbol{y}_i, \boldsymbol{\delta}^{*(r)}\big) = \frac{f\big(\boldsymbol{y}_i|b_i^*, \boldsymbol{\delta}^{*(r)}\big)\phi_2(b_i^*)}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f\big(\boldsymbol{y}_i|b_i^*, \boldsymbol{\delta}^{*(r)}\big)\phi_2(b_i^*) dc_i^* dd_i^*}.$$

We can utilize the method of GHQ to approximate the integrals. Let $q_{l_1}^c$ and $q_{l_2}^d$ are abscissas of m-point GHQ corresponding to the random effects of component-1 and component-2, respectively with associated respective weights $w_{l_1}^c$ and $w_{l_2}^d; l_1, l_2 = 1, \cdots, m$. Let us define $g_i^{(r)} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f\big(\boldsymbol{y}_i|b_i^*, \boldsymbol{\delta}^{*(r)}\big)\phi_2(b_i^*) dc_i^* dd_i^*$. Then

$$g_i^{(r)} \approx \sum_{l_1=1}^{m}\sum_{l_2=1}^{m} \frac{1}{\pi} f\big(\boldsymbol{y}_i|q_{l_1}^{*c}, q_{l_2}^{*d}, \boldsymbol{\delta}^{*(r)}\big) w_{l_1}^c w_{l_2}^d = \sum_{l_1=1}^{m}\sum_{l_2=1}^{m} f\big(\boldsymbol{y}_i|q_{l_1}^{*c}, q_{l_2}^{*d}, \boldsymbol{\delta}^{*(r)}\big) w_{l_1}^{*c} w_{l_2}^{*d},$$

where $q^{*c}_{l_1} = \sqrt{2} q^c_{l_1}, q^{*d}_{l_2} = \sqrt{2} q^d_{l_2}, w^{*c}_{l_1} = w^c_{l_1}/\sqrt{\pi}, w^{*d}_{l_2} = w^d_{l_2}/\sqrt{\pi}$. Then, we have

$$
\begin{aligned}
Q(\boldsymbol{\delta}^*|\boldsymbol{\delta}^{*(r)}) \approx & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m} w^*_{il_1l_2}\Big[u^{(r)}_{ij}(q^{*c}_{l_1}, q^{*d}_{l_2})\log(\pi^*) + \big(1-u^{(r)}_{ij}(q^{*c}_{l_1}, q^{*d}_{l_2})\big)\log(1-\pi^*)\Big] + \\
& \sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m} w^*_{il_1l_2} u^{(r)}_{ij}(q^{*c}_{l_1}, q^{*d}_{l_2})\Big[y_{ij}\big(\boldsymbol{x}'_{ij}\boldsymbol{\alpha} + \sigma_c q^{*c}_{l_1}\big) - e^{\boldsymbol{x}'_{ij}\boldsymbol{\alpha}+\sigma_c q^{*c}_{l_1}}\Big] + \\
& \sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m} w^*_{il_1l_2}\big(1-u^{(r)}_{ij}(q^{*c}_{l_1}, q^{*d}_{l_2})\big)\Big[y_{ij}\big(\boldsymbol{x}'_{ij}\boldsymbol{\gamma} + \sigma_d q^{*d}_{l_2}\big) - e^{\boldsymbol{x}'_{ij}\boldsymbol{\gamma}+\sigma_d q^{*d}_{l_2}}\Big] \\
= & \ell_{\pi^*} + \ell_{\boldsymbol{\alpha}',\sigma_c} + \ell_{\boldsymbol{\gamma}',\sigma_d},
\end{aligned}
\tag{B.14}
$$

where $w^*_{il_1l_2} = f\big(\boldsymbol{y}_i|q^{*c}_{l_1}, q^{*d}_{l_2}, \boldsymbol{\delta}^{*(r)}\big) w^{*c}_{l_1} w^{*d}_{l_2}/g^{(r)}_i = \prod_{j=1}^{n_i} f\big(y_{ij}|q^{*c}_{l_1}, q^{*d}_{l_2}, \boldsymbol{\delta}^{*(r)}\big) w^{*c}_{l_1} w^{*d}_{l_2}/g^{(r)}_i$ (constant over index j).

M-Step:

Like EM algorithm of Poisson-Poisson mixture model for cross-sectional setup, from Eq.(B.14) we observed that $Q(\boldsymbol{\delta}^*|\boldsymbol{\delta}^{*(r)})$ can be decomposed into three components $\ell_{\pi^*}, \ell_{\boldsymbol{\alpha}',\sigma_c}$ and $\ell_{\boldsymbol{\gamma}',\sigma_d}$. Therefore, these can be optimized separately.

M-Step for π^*:

At the $(r+1)$th step, maximization of ℓ_{π^*} of Eq.(B.14) with respect to π^* gives the next updated estimate of π^* as

$$
\pi^{*(r+1)} = \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m} w^*_{il_1l_2} u^{(r)}_{ij}(q^{*c}_{l_1}, q^{*d}_{l_2})}{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m} w^*_{il_1l_2}}. \tag{B.15}
$$

M-Step for $\boldsymbol{\alpha}, \sigma_c$:

Maximization of $\ell_{\boldsymbol{\alpha}',\sigma_c}$ of Eq.(B.14) with respect to $\boldsymbol{\alpha}, \sigma_c$ can be done simultaneously by fitting a weighted Poisson regression model. To do this, let us define $\boldsymbol{x}^* = \big[\big(\boldsymbol{x}\otimes\mathbf{1}_{m^2}\big), \big\{\mathbf{1}_n \otimes \big(\mathbf{1}_m \otimes (q^{*c}_1, \ldots, q^{*c}_m)'\big)\big\}\big]_{nm^2\times(p+1)}, \boldsymbol{\alpha}^*_{(p+1)\times 1} = (\boldsymbol{\alpha}', \sigma_c)'$. Here $\mathbf{1}_h$ is the $h\times 1$ vector of 1s.

Hence maximization with respect to $\boldsymbol{\alpha}^*$ can be accomplished by fitting a weighted log-linear regression of $\boldsymbol{y} \otimes \mathbf{1}_{m^2}$ on $\boldsymbol{x}^*$ with weights $w^*_{il_1l_2} u^{(r)}_{ij}(q^{*c}_{l_1}, q^{*d}_{l_2}), i = 1, \cdots, K, j = 1, \cdots, n_i, l_1 = l_2 = 1, \cdots, m$.

M-Step for $\boldsymbol{\gamma}, \sigma_d$:

Maximization of $\ell_{\boldsymbol{\gamma}', \sigma_d}$ of Eq.(B.14) with respect to $\boldsymbol{\gamma}, \sigma_d$ can be done simultaneously by fitting another weighted Poisson regression model. To do this, let us define $\boldsymbol{x}^* = \big[\big(\boldsymbol{x} \otimes \mathbf{1}_{m^2}\big), \big\{\mathbf{1}_n \otimes \big((q^{*d}_1, \ldots, q^{*d}_m)' \otimes \mathbf{1}_m\big)\big\}\big]_{nm^2 \times (p+1)}, \boldsymbol{\gamma}^*_{(p+1)\times 1} = (\boldsymbol{\gamma}', \sigma_d)'$. Hence maximization with respect to $\boldsymbol{\gamma}^*$ can be accomplished by fitting a weighted log-linear regression of $\boldsymbol{y} \otimes \mathbf{1}_{m^2}$ on $\boldsymbol{x}^*$ with weights $w^*_{il_1l_2}\big(1 - u^{(r)}_{ij}(q^{*c}_{l_1}, q^{*d}_{l_2})\big), i = 1, \cdots, K, j = 1, \cdots, n_i, l_1 = l_2 = 1, \cdots, m$.

B.4 Simulation Results: Intra-cluster Correlation Coefficient from the REMPois-Pois Model

Table B.1: The ICC and the estimated ICC from REMPois-Pois model for different number of clusters and various cluster sizes

π^*	K	n_i	ICC	$\hat{\text{ICC}}$
0.50	50	5	0.108	0.098
		15	0.107	0.102
		30	0.108	0.106
	100	5	0.109	0.100
		15	0.108	0.108
		30	0.108	0.104
	200	5	0.108	0.104
		15	0.107	0.105
		30	0.108	0.102
0.70	50	5	0.062	0.062
		15	0.062	0.065
		30	0.062	0.062
	100	5	0.063	0.061
		15	0.062	0.062

Continued...Table B.1

π^*	K	n_i	ICC	$\hat{\text{ICC}}$
		30	0.062	0.060
	200	5	0.062	0.061
		15	0.062	0.061
		30	0.062	0.058
0.90	50	5	0.020	0.037
		15	0.020	0.022
		30	0.020	0.020
	100	5	0.020	0.024
		15	0.020	0.020
		30	0.020	0.018
	200	5	0.020	0.021
		15	0.020	0.019
		30	0.020	0.017

Appendix C

C.1 Simulation Results: Marginalized Mixture of Poisson AR(1) Model

Table C.1: Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.50$ from marginalized mixture of Poisson AR(1) model for $\rho^* = 0.40$ and for different values of K and T

(K, T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(100,3)		$\beta_0 = 0.20$	0.196	-0.004	0.225	0.196	95.6
		$\beta_1 = 0.60$	0.584	-0.016	0.340	0.293	95.6
		$\beta_2 = 0.50$	0.493	-0.007	0.237	0.216	95.5
		$\rho_1 = 0.691$	0.685	-0.006			
		$\alpha_0 = -1.00$	-1.192	-0.192	0.349	0.706	94.3
		$\alpha_1 = 0.40$	0.545	0.145	0.510	0.878	95.1
	0.907	$\alpha_2 = 0.50$	0.586	0.086	0.350	0.588	94.6
		$\rho_1^* = 0.40$	0.175	-0.225			
		$\pi^* = 0.50$	0.500	-0.000			
(100,4)		$\beta_0 = 0.20$	0.184	-0.016	0.217	0.175	95.3
		$\beta_1 = 0.60$	0.605	0.005	0.328	0.279	95.9
		$\beta_2 = 0.50$	0.503	0.003	0.228	0.202	96.1
		$\rho_1 = 0.691$	0.685	-0.006			
		$\alpha_0 = -1.00$	-1.111	-0.111	0.310	0.604	95.8
		$\alpha_1 = 0.40$	0.485	0.085	0.458	0.801	96.2
	0.899	$\alpha_2 = 0.50$	0.581	0.081	0.310	0.509	94.8
		$\rho_1^* = 0.40$	0.251	-0.149			
		$\pi^* = 0.50$	0.504	0.004			
(100,5)		$\beta_0 = 0.20$	0.192	-0.008	0.210	0.167	95.5
		$\beta_1 = 0.60$	0.591	-0.009	0.318	0.263	96.4
		$\beta_2 = 0.50$	0.517	0.017	0.221	0.194	95.7
		$\rho_1 = 0.691$	0.688	-0.003			
		$\alpha_0 = -1.00$	-1.077	-0.077	0.288	0.529	95.4
		$\alpha_1 = 0.40$	0.448	0.048	0.426	0.685	96.4
	0.870	$\alpha_2 = 0.50$	0.555	0.055	0.288	0.461	95.3

Continued...Table C.1

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
		$\rho_1^* = 0.40$	0.294	-0.106			
		$\pi^* = 0.50$	0.496	-0.004			
(200,3)		$\beta_0 = 0.20$	0.194	-0.006	0.161	0.136	95.9
		$\beta_1 = 0.60$	0.599	-0.001	0.245	0.214	95.9
		$\beta_2 = 0.50$	0.499	-0.001	0.159	0.142	95.5
		$\rho_1 = 0.691$	0.690	-0.001			
		$\alpha_0 = -1.00$	-1.159	-0.159	0.246	0.450	93.2
		$\alpha_1 = 0.40$	0.546	0.146	0.364	0.570	94.3
	0.913	$\alpha_2 = 0.50$	0.578	0.078	0.228	0.343	95.0
		$\rho_1^* = 0.40$	0.187	-0.213			
		$\pi^* = 0.50$	0.498	-0.002			
(200,4)		$\beta_0 = 0.20$	0.196	-0.004	0.155	0.133	96.6
		$\beta_1 = 0.60$	0.594	-0.006	0.237	0.206	94.8
		$\beta_2 = 0.50$	0.497	-0.003	0.153	0.141	95.6
		$\rho_1 = 0.691$	0.694	0.003			
		$\alpha_0 = -1.00$	-1.066	-0.066	0.219	0.372	94.8
		$\alpha_1 = 0.40$	0.458	0.058	0.328	0.475	95.1
	0.903	$\alpha_2 = 0.50$	0.545	0.045	0.205	0.300	95.2
		$\rho_1^* = 0.40$	0.264	-0.136			
		$\pi^* = 0.50$	0.504	0.004			
(200,5)		$\beta_0 = 0.20$	0.191	-0.009	0.151	0.127	96.4
		$\beta_1 = 0.60$	0.598	-0.002	0.230	0.204	96.6
		$\beta_2 = 0.50$	0.513	0.013	0.148	0.132	95.7
		$\rho_1 = 0.691$	0.693	0.002			
		$\alpha_0 = -1.00$	-1.056	-0.056	0.205	0.337	94.5
		$\alpha_1 = 0.40$	0.452	0.052	0.307	0.455	94.5
	0.911	$\alpha_2 = 0.50$	0.543	0.043	0.191	0.274	95.2
		$\rho_1^* = 0.40$	0.304	-0.096			
		$\pi^* = 0.50$	0.498	-0.002			
(500,3)		$\beta_0 = 0.20$	0.201	0.001	0.097	0.081	96.2
		$\beta_1 = 0.60$	0.593	-0.007	0.160	0.136	96.1
		$\beta_2 = 0.50$	0.499	-0.001	0.102	0.090	95.6
		$\rho_1 = 0.691$	0.690	-0.001			
		$\alpha_0 = -1.00$	-1.129	-0.129	0.145	0.272	91.7
		$\alpha_1 = 0.40$	0.516	0.116	0.232	0.363	93.7
	0.914	$\alpha_2 = 0.50$	0.568	0.068	0.142	0.215	93.3
		$\rho_1^* = 0.40$	0.189	-0.211			
		$\pi^* = 0.50$	0.498	-0.002			
(500,4)		$\beta_0 = 0.20$	0.202	0.002	0.093	0.074	95.5
		$\beta_1 = 0.60$	0.593	-0.007	0.154	0.128	96.0
		$\beta_2 = 0.50$	0.496	-0.004	0.098	0.086	95.9
		$\rho_1 = 0.691$	0.690	-0.001			
		$\alpha_0 = -1.00$	-1.063	-0.063	0.131	0.225	94.0
		$\alpha_1 = 0.40$	0.469	0.069	0.211	0.308	93.9
	0.881	$\alpha_2 = 0.50$	0.540	0.040	0.129	0.190	94.0
		$\rho_1^* = 0.40$	0.262	-0.138			
		$\pi^* = 0.50$	0.501	0.001			
(500,5)		$\beta_0 = 0.20$	0.197	-0.003	0.090	0.075	95.6

Continued...Table C.1

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
		$\beta_1 = 0.60$	0.603	0.003	0.149	0.129	96.6
		$\beta_2 = 0.50$	0.500	-0.000	0.095	0.085	95.7
		$\rho_1 = 0.691$	0.689	-0.002			
		$\alpha_0 = -1.00$	-1.028	-0.028	0.122	0.199	95.5
		$\alpha_1 = 0.40$	0.437	0.037	0.197	0.266	94.0
	0.915	$\alpha_2 = 0.50$	0.518	0.018	0.121	0.165	95.5
		$\rho_1^* = 0.40$	0.307	-0.093			
		$\pi^* = 0.50$	0.500	-0.000			

Table C.2: Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.90$ from marginalized mixture of Poisson AR(1) model for $\rho^* = 0.40$ and for different values of K and T

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(100,3)		$\beta_0 = 0.20$	0.195	-0.005	0.507	0.569	95.6
		$\beta_1 = 0.60$	0.557	-0.043	0.792	0.938	94.4
		$\beta_2 = 0.50$	0.398	-0.102	0.555	0.687	95.5
		$\rho_1 = 0.935$	0.908	-0.027			
		$\alpha_0 = -1.00$	-1.015	-0.015	0.278	0.286	93.9
		$\alpha_1 = 0.40$	0.389	-0.011	0.413	0.429	95.0
	0.825	$\alpha_2 = 0.50$	0.490	-0.010	0.280	0.301	94.4
		$\rho_1^* = 0.40$	0.380	-0.020			
		$\pi^* = 0.90$	0.896	-0.004			
(100,4)		$\beta_0 = 0.20$	0.195	-0.005	0.500	0.577	95.2
		$\beta_1 = 0.60$	0.551	-0.049	0.781	0.934	94.4
		$\beta_2 = 0.50$	0.404	-0.096	0.552	0.700	95.2
		$\rho_1 = 0.935$	0.906	-0.029			
		$\alpha_0 = -1.00$	-1.004	-0.004	0.248	0.251	94.5
		$\alpha_1 = 0.40$	0.392	-0.008	0.369	0.383	94.7
	0.833	$\alpha_2 = 0.50$	0.483	-0.017	0.250	0.263	94.4
		$\rho_1^* = 0.40$	0.384	-0.016			
		$\pi^* = 0.90$	0.896	-0.004			
(100,5)		$\beta_0 = 0.20$	0.207	0.007	0.497	0.564	94.6
		$\beta_1 = 0.60$	0.519	-0.081	0.774	0.922	94.5
		$\beta_2 = 0.50$	0.404	-0.096	0.547	0.680	94.0
		$\rho_1 = 0.935$	0.907	-0.028			
		$\alpha_0 = -1.00$	-1.004	-0.004	0.227	0.224	95.0
		$\alpha_1 = 0.40$	0.401	0.001	0.337	0.339	95.6
	0.815	$\alpha_2 = 0.50$	0.488	-0.012	0.229	0.228	95.3
		$\rho_1^* = 0.40$	0.385	-0.015			
		$\pi^* = 0.90$	0.896	-0.004			
(200,3)		$\beta_0 = 0.20$	0.158	-0.042	0.371	0.367	95.2
		$\beta_1 = 0.60$	0.633	0.033	0.590	0.568	94.9
		$\beta_2 = 0.50$	0.500	-0.000	0.386	0.362	95.4
		$\rho_1 = 0.935$	0.931	-0.004			

Continued...Table C.2

(K, T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
		$\alpha_0 = -1.00$	-1.020	-0.020	0.199	0.205	94.8
		$\alpha_1 = 0.40$	0.420	0.020	0.298	0.307	94.9
	0.811	$\alpha_2 = 0.50$	0.511	0.011	0.186	0.191	94.9
		$\rho_1^* = 0.40$	0.390	-0.010			
		$\pi^* = 0.90$	0.900	0.000			
(200,4)		$\beta_0 = 0.20$	0.154	-0.046	0.366	0.363	94.4
		$\beta_1 = 0.60$	0.622	0.022	0.581	0.585	95.0
		$\beta_2 = 0.50$	0.479	-0.021	0.380	0.383	95.0
		$\rho_1 = 0.935$	0.930	-0.005			
		$\alpha_0 = -1.00$	-1.011	-0.011	0.179	0.189	95.3
		$\alpha_1 = 0.40$	0.409	0.009	0.268	0.278	94.4
	0.854	$\alpha_2 = 0.50$	0.499	-0.001	0.167	0.174	95.1
		$\rho_1^* = 0.40$	0.392	-0.008			
		$\pi^* = 0.90$	0.900	-0.000			
(200,5)		$\beta_0 = 0.20$	0.170	-0.030	0.363	0.348	95.3
		$\beta_1 = 0.60$	0.606	0.006	0.577	0.570	95.0
		$\beta_2 = 0.50$	0.522	0.022	0.379	0.347	95.3
		$\rho_1 = 0.935$	0.932	-0.003			
		$\alpha_0 = -1.00$	-1.007	-0.007	0.164	0.172	95.9
		$\alpha_1 = 0.40$	0.400	-0.000	0.246	0.262	95.3
	0.812	$\alpha_2 = 0.50$	0.503	0.003	0.153	0.153	95.2
		$\rho_1^* = 0.40$	0.391	-0.009			
		$\pi^* = 0.90$	0.898	-0.002			
(500,3)		$\beta_0 = 0.20$	0.197	-0.003	0.223	0.193	95.0
		$\beta_1 = 0.60$	0.586	-0.014	0.385	0.337	95.0
		$\beta_2 = 0.50$	0.481	-0.019	0.250	0.228	95.4
		$\rho_1 = 0.935$	0.932	-0.003			
		$\alpha_0 = -1.00$	-0.997	0.003	0.119	0.117	94.6
		$\alpha_1 = 0.40$	0.390	-0.010	0.193	0.192	95.1
	0.863	$\alpha_2 = 0.50$	0.492	-0.008	0.118	0.117	94.7
		$\rho_1^* = 0.40$	0.396	-0.004			
		$\pi^* = 0.90$	0.899	-0.001			
(500,4)		$\beta_0 = 0.20$	0.196	-0.004	0.223	0.203	96.3
		$\beta_1 = 0.60$	0.576	-0.024	0.382	0.368	96.3
		$\beta_2 = 0.50$	0.489	-0.011	0.248	0.230	96.0
		$\rho_1 = 0.935$	0.934	-0.001			
		$\alpha_0 = -1.00$	-0.998	0.002	0.107	0.109	94.5
		$\alpha_1 = 0.40$	0.393	-0.007	0.174	0.173	95.2
	0.901	$\alpha_2 = 0.50$	0.500	-0.000	0.107	0.102	94.7
		$\rho_1^* = 0.40$	0.398	-0.002			
		$\pi^* = 0.90$	0.900	-0.000			
(500,5)		$\beta_0 = 0.20$	0.178	-0.022	0.222	0.187	95.1
		$\beta_1 = 0.60$	0.603	0.003	0.382	0.331	96.2
		$\beta_2 = 0.50$	0.495	-0.005	0.248	0.212	95.7
		$\rho_1 = 0.935$	0.933	-0.002			
		$\alpha_0 = -1.00$	-1.004	-0.004	0.099	0.096	95.1
		$\alpha_1 = 0.40$	0.403	0.003	0.160	0.151	94.8
	0.869	$\alpha_2 = 0.50$	0.499	-0.001	0.098	0.099	95.2
		$\rho_1^* = 0.40$	0.397	-0.003			

Continued...Table C.2

(K, T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
		$\pi^* = 0.90$	0.900	0.000			

C.2 Simulation Results: Marginalized Mixture of Poisson MA(1) Model

Table C.3: Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.50$ from marginalized mixture of Poisson MA(1) model for $\rho^* = 0.40$ and for different values of K and T

(K, T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(100,3)		$\beta_0 = 0.20$	0.204	0.004	0.216	0.199	95.8
		$\beta_1 = 0.60$	0.590	-0.010	0.326	0.308	95.2
		$\beta_2 = 0.50$	0.496	-0.004	0.228	0.207	96.4
		$\rho_1 = 0.632$	0.626	-0.006			
		$\alpha_0 = -1.00$	-1.114	-0.114	0.325	0.648	95.8
		$\alpha_1 = 0.40$	0.496	0.096	0.479	0.782	95.2
	0.924	$\alpha_2 = 0.50$	0.558	0.058	0.326	0.537	95.2
		$\rho_1^* = 0.286$	0.121	-0.165			
		$\pi^* = 0.50$	0.496	-0.004			
(100,4)		$\beta_0 = 0.20$	0.190	-0.010	0.208	0.181	96.1
		$\beta_1 = 0.60$	0.607	0.007	0.314	0.281	96.4
		$\beta_2 = 0.50$	0.500	0.000	0.220	0.201	96.1
		$\rho_1 = 0.632$	0.628	-0.004			
		$\alpha_0 = -1.00$	-1.076	-0.076	0.289	0.532	95.2
		$\alpha_1 = 0.40$	0.468	0.068	0.426	0.685	95.2
	0.915	$\alpha_2 = 0.50$	0.544	0.044	0.289	0.447	95.2
		$\rho_1^* = 0.286$	0.188	-0.098			
		$\pi^* = 0.50$	0.494	-0.006			
(100,5)		$\beta_0 = 0.20$	0.186	-0.014	0.203	0.178	96.1
		$\beta_1 = 0.60$	0.619	0.019	0.307	0.278	96.5
		$\beta_2 = 0.50$	0.509	0.009	0.214	0.181	95.8
		$\rho_1 = 0.632$	0.629	-0.003			
		$\alpha_0 = -1.00$	-1.024	-0.024	0.261	0.442	95.7
		$\alpha_1 = 0.40$	0.450	0.050	0.387	0.591	95.4
	0.891	$\alpha_2 = 0.50$	0.512	0.012	0.264	0.397	95.2
		$\rho_1^* = 0.286$	0.237	-0.048			
		$\pi^* = 0.50$	0.497	-0.003			
(200,3)		$\beta_0 = 0.20$	0.202	0.002	0.155	0.125	96.1
		$\beta_1 = 0.60$	0.591	-0.009	0.237	0.199	95.7
		$\beta_2 = 0.50$	0.504	0.004	0.153	0.124	96.2
		$\rho_1 = 0.632$	0.636	0.004			

Continued...Table C.3

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
		$\alpha_0 = -1.00$	-1.080	-0.080	0.229	0.392	94.6
		$\alpha_1 = 0.40$	0.472	0.072	0.341	0.495	94.5
	0.894	$\alpha_2 = 0.50$	0.553	0.053	0.213	0.309	94.4
		$\rho_1^* = 0.286$	0.138	-0.148			
		$\pi^* = 0.50$	0.496	-0.004			
(200,4)		$\beta_0 = 0.20$	0.190	-0.010	0.150	0.121	95.5
		$\beta_1 = 0.60$	0.609	0.009	0.228	0.192	95.2
		$\beta_2 = 0.50$	0.497	-0.003	0.148	0.127	95.3
		$\rho_1 = 0.632$	0.635	0.003			
		$\alpha_0 = -1.00$	-1.030	-0.030	0.204	0.330	94.6
		$\alpha_1 = 0.40$	0.449	0.049	0.305	0.414	94.8
	0.890	$\alpha_2 = 0.50$	0.519	0.019	0.191	0.272	95.3
		$\rho_1^* = 0.286$	0.209	-0.077			
		$\pi^* = 0.50$	0.501	0.001			
(200,5)		$\beta_0 = 0.20$	0.199	-0.001	0.145	0.125	96.5
		$\beta_1 = 0.60$	0.598	-0.002	0.222	0.195	96.6
		$\beta_2 = 0.50$	0.498	-0.002	0.143	0.132	96.3
		$\rho_1 = 0.632$	0.635	0.003			
		$\alpha_0 = -1.00$	-0.990	0.010	0.186	0.295	94.5
		$\alpha_1 = 0.40$	0.393	-0.007	0.280	0.383	94.3
	0.916	$\alpha_2 = 0.50$	0.502	0.002	0.175	0.241	95.2
		$\rho_1^* = 0.286$	0.246	-0.039			
		$\pi^* = 0.50$	0.497	-0.003			
(500,3)		$\beta_0 = 0.20$	0.197	-0.003	0.093	0.078	96.8
		$\beta_1 = 0.60$	0.601	0.001	0.154	0.134	95.7
		$\beta_2 = 0.50$	0.499	-0.001	0.098	0.088	95.7
		$\rho_1 = 0.632$	0.630	-0.002			
		$\alpha_0 = -1.00$	-1.075	-0.075	0.135	0.249	93.3
		$\alpha_1 = 0.40$	0.500	0.100	0.217	0.319	93.9
	0.899	$\alpha_2 = 0.50$	0.542	0.042	0.133	0.206	94.1
		$\rho_1^* = 0.286$	0.134	-0.152			
		$\pi^* = 0.50$	0.499	-0.001			
(500,4)		$\beta_0 = 0.20$	0.198	-0.002	0.089	0.071	94.7
		$\beta_1 = 0.60$	0.601	0.001	0.148	0.125	96.1
		$\beta_2 = 0.50$	0.497	-0.003	0.094	0.081	95.9
		$\rho_1 = 0.632$	0.631	-0.001			
		$\alpha_0 = -1.00$	-1.021	-0.021	0.121	0.195	95.6
		$\alpha_1 = 0.40$	0.434	0.034	0.196	0.275	95.2
	0.894	$\alpha_2 = 0.50$	0.513	0.013	0.120	0.165	95.1
		$\rho_1^* = 0.286$	0.204	-0.082			
		$\pi^* = 0.50$	0.499	-0.001			
(500,5)		$\beta_0 = 0.20$	0.200	-0.000	0.087	0.069	95.7
		$\beta_1 = 0.60$	0.592	-0.008	0.144	0.120	95.4
		$\beta_2 = 0.50$	0.503	0.003	0.092	0.081	96.1
		$\rho_1 = 0.632$	0.630	-0.002			
		$\alpha_0 = -1.00$	-0.985	0.015	0.111	0.171	95.4
		$\alpha_1 = 0.40$	0.405	0.005	0.180	0.234	95.6
	0.900	$\alpha_2 = 0.50$	0.501	0.001	0.110	0.147	94.7
		$\rho_1^* = 0.286$	0.248	-0.038			

Continued...Table C.3

(K, T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
		$\pi^* = 0.50$	0.500	0.000			

Table C.4: Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.90$ from marginalized mixture of Poisson MA(1) model for $\rho^* = 0.40$ and for different values of K and T

(K, T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(100,3)		$\beta_0 = 0.20$	0.202	0.002	0.508	0.533	94.6
		$\beta_1 = 0.60$	0.539	-0.061	0.790	0.879	95.1
		$\beta_2 = 0.50$	0.426	-0.074	0.559	0.654	94.0
		$\rho_1 = 0.923$	0.895	-0.028			
		$\alpha_0 = -1.00$	-1.027	-0.027	0.255	0.253	94.2
		$\alpha_1 = 0.40$	0.411	0.011	0.378	0.376	95.0
	0.834	$\alpha_2 = 0.50$	0.507	0.007	0.256	0.254	95.6
		$\rho_1^* = 0.286$	0.269	-0.017			
		$\pi^* = 0.90$	0.896	-0.004			
(100,4)		$\beta_0 = 0.20$	0.165	-0.035	0.497	0.575	94.4
		$\beta_1 = 0.60$	0.585	-0.015	0.777	0.961	94.6
		$\beta_2 = 0.50$	0.436	-0.064	0.547	0.623	93.9
		$\rho_1 = 0.923$	0.891	-0.032			
		$\alpha_0 = -1.00$	-0.999	0.001	0.222	0.229	95.2
		$\alpha_1 = 0.40$	0.388	-0.012	0.331	0.335	95.4
	0.821	$\alpha_2 = 0.50$	0.489	-0.011	0.224	0.234	95.2
		$\rho_1^* = 0.286$	0.275	-0.011			
		$\pi^* = 0.90$	0.897	-0.003			
(100,5)		$\beta_0 = 0.20$	0.201	0.001	0.497	0.598	95.1
		$\beta_1 = 0.60$	0.528	-0.072	0.775	0.993	94.0
		$\beta_2 = 0.50$	0.378	-0.122	0.544	0.698	93.6
		$\rho_1 = 0.923$	0.890	-0.033			
		$\alpha_0 = -1.00$	-1.014	-0.014	0.202	0.204	94.5
		$\alpha_1 = 0.40$	0.402	0.002	0.299	0.305	94.9
	0.823	$\alpha_2 = 0.50$	0.506	0.006	0.202	0.208	94.9
		$\rho_1^* = 0.286$	0.274	-0.012			
		$\pi^* = 0.90$	0.897	-0.003			
(200,3)		$\beta_0 = 0.20$	0.206	0.006	0.368	0.358	94.2
		$\beta_1 = 0.60$	0.577	-0.023	0.582	0.586	95.1
		$\beta_2 = 0.50$	0.454	-0.046	0.379	0.388	94.7
		$\rho_1 = 0.923$	0.917	-0.006			
		$\alpha_0 = -1.00$	-1.008	-0.008	0.182	0.186	95.1
		$\alpha_1 = 0.40$	0.404	0.004	0.273	0.283	95.3
	0.850	$\alpha_2 = 0.50$	0.499	-0.001	0.171	0.171	95.9
		$\rho_1^* = 0.286$	0.278	-0.008			
		$\pi^* = 0.90$	0.898	-0.002			
(200,4)		$\beta_0 = 0.20$	0.160	-0.040	0.363	0.327	94.8

Continued...Table C.4

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
		$\beta_1 = 0.60$	0.607	0.007	0.574	0.574	95.7
		$\beta_2 = 0.50$	0.516	0.016	0.378	0.376	94.9
		$\rho_1 = 0.923$	0.919	-0.004			
		$\alpha_0 = -1.00$	-1.001	-0.001	0.160	0.164	94.4
		$\alpha_1 = 0.40$	0.401	0.001	0.240	0.246	95.3
	0.851	$\alpha_2 = 0.50$	0.494	-0.006	0.150	0.155	95.1
		$\rho_1^* = 0.286$	0.284	-0.002			
		$\pi^* = 0.90$	0.899	-0.001			
(200,5)		$\beta_0 = 0.20$	0.184	-0.016	0.362	0.350	95.2
		$\beta_1 = 0.60$	0.592	-0.008	0.574	0.582	94.9
		$\beta_2 = 0.50$	0.503	0.003	0.376	0.358	95.3
		$\rho_1 = 0.923$	0.920	-0.003			
		$\alpha_0 = -1.00$	-1.002	-0.002	0.144	0.148	94.1
		$\alpha_1 = 0.40$	0.394	-0.006	0.217	0.227	94.3
	0.846	$\alpha_2 = 0.50$	0.504	0.004	0.135	0.138	94.4
		$\rho_1^* = 0.286$	0.279	-0.006			
		$\pi^* = 0.90$	0.898	-0.002			
(500,3)		$\beta_0 = 0.20$	0.179	-0.021	0.222	0.194	95.4
		$\beta_1 = 0.60$	0.620	0.020	0.382	0.343	96.2
		$\beta_2 = 0.50$	0.517	0.017	0.249	0.223	95.7
		$\rho_1 = 0.923$	0.920	-0.003			
		$\alpha_0 = -1.00$	-1.003	-0.003	0.109	0.112	95.4
		$\alpha_1 = 0.40$	0.405	0.005	0.177	0.178	95.3
	0.876	$\alpha_2 = 0.50$	0.502	0.002	0.108	0.112	94.4
		$\rho_1^* = 0.286$	0.283	-0.003			
		$\pi^* = 0.90$	0.899	-0.001			
(500,4)		$\beta_0 = 0.20$	0.197	-0.003	0.222	0.187	96.1
		$\beta_1 = 0.60$	0.587	-0.013	0.381	0.320	94.4
		$\beta_2 = 0.50$	0.485	-0.015	0.247	0.217	96.5
		$\rho_1 = 0.923$	0.921	-0.002			
		$\alpha_0 = -1.00$	-1.009	-0.009	0.096	0.095	94.8
		$\alpha_1 = 0.40$	0.411	0.011	0.156	0.152	94.9
	0.869	$\alpha_2 = 0.50$	0.502	0.002	0.096	0.094	95.3
		$\rho_1^* = 0.286$	0.283	-0.002			
		$\pi^* = 0.90$	0.899	-0.001			
(500,5)		$\beta_0 = 0.20$	0.185	-0.015	0.221	0.182	94.6
		$\beta_1 = 0.60$	0.595	-0.005	0.379	0.324	96.4
		$\beta_2 = 0.50$	0.500	0.000	0.246	0.225	95.4
		$\rho_1 = 0.923$	0.922	-0.001			
		$\alpha_0 = -1.00$	-1.002	-0.002	0.087	0.086	95.5
		$\alpha_1 = 0.40$	0.400	0.000	0.141	0.140	95.4
	0.883	$\alpha_2 = 0.50$	0.507	0.007	0.086	0.082	95.0
		$\rho_1^* = 0.286$	0.284	-0.002			
		$\pi^* = 0.90$	0.900	-0.000			

C.3 Simulation Results: Marginalized Mixture of Poisson EQCOR Model

Table C.5: Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.50$ from marginalized mixture of Poisson EQCOR model for $\rho^* = 0.40$ and for different values of K and T

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(100,3)		$\beta_0 = 0.20$	0.189	-0.011	0.217	0.185	95.6
		$\beta_1 = 0.60$	0.595	-0.005	0.328	0.282	95.4
		$\beta_2 = 0.50$	0.498	-0.002	0.229	0.207	95.5
		$\rho_1 = 0.567$	0.565	-0.002			
		$\alpha_0 = -1.00$	-1.102	-0.102	0.314	0.633	96.3
		$\alpha_1 = 0.40$	0.499	0.099	0.462	0.772	95.9
	0.895	$\alpha_2 = 0.50$	0.573	0.073	0.314	0.533	95.6
		$\rho_1^* = 0.16$	0.039	-0.121			
		$\pi^* = 0.50$	0.505	0.005			
(100,4)		$\beta_0 = 0.20$	0.191	-0.009	0.210	0.177	95.7
		$\beta_1 = 0.60$	0.604	0.004	0.318	0.277	96.2
		$\beta_2 = 0.50$	0.487	-0.013	0.222	0.208	95.1
		$\rho_1 = 0.567$	0.561	-0.006			
		$\alpha_0 = -1.00$	-1.072	-0.072	0.288	0.553	95.5
		$\alpha_1 = 0.40$	0.470	0.070	0.425	0.679	95.2
	0.903	$\alpha_2 = 0.50$	0.540	0.040	0.289	0.481	94.8
		$\rho_1^* = 0.16$	0.073	-0.087			
		$\pi^* = 0.50$	0.500	-0.000			
(100,5)		$\beta_0 = 0.20$	0.205	0.005	0.206	0.172	95.8
		$\beta_1 = 0.60$	0.584	-0.016	0.312	0.274	95.1
		$\beta_2 = 0.50$	0.487	-0.013	0.218	0.195	95.6
		$\rho_1 = 0.567$	0.562	-0.005			
		$\alpha_0 = -1.00$	-1.015	-0.015	0.268	0.474	94.9
		$\alpha_1 = 0.40$	0.400	0.000	0.399	0.619	95.7
	0.901	$\alpha_2 = 0.50$	0.516	0.016	0.271	0.398	94.6
		$\rho_1^* = 0.16$	0.093	-0.067			
		$\pi^* = 0.50$	0.498	-0.002			
(200,3)		$\beta_0 = 0.20$	0.188	-0.012	0.155	0.134	96.0
		$\beta_1 = 0.60$	0.607	0.007	0.236	0.210	96.7
		$\beta_2 = 0.50$	0.506	0.006	0.153	0.132	96.0
		$\rho_1 = 0.567$	0.571	0.004			
		$\alpha_0 = -1.00$	-1.081	-0.081	0.222	0.397	94.6
		$\alpha_1 = 0.40$	0.504	0.104	0.330	0.490	94.5
	0.907	$\alpha_2 = 0.50$	0.536	0.036	0.207	0.307	95.0
		$\rho_1^* = 0.16$	0.042	-0.118			
		$\pi^* = 0.50$	0.499	-0.001			
(200,4)		$\beta_0 = 0.20$	0.202	0.002	0.151	0.130	96.8
		$\beta_1 = 0.60$	0.589	-0.011	0.231	0.205	95.9
		$\beta_2 = 0.50$	0.493	-0.007	0.149	0.123	94.6

Continued...Table C.5

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
		$\rho_1 = 0.567$	0.572	0.005			
		$\alpha_0 = -1.00$	-1.046	-0.046	0.205	0.342	95.1
		$\alpha_1 = 0.40$	0.442	0.042	0.307	0.445	95.6
	0.903	$\alpha_2 = 0.50$	0.530	0.030	0.192	0.277	95.1
		$\rho_1^* = 0.16$	0.088	-0.072			
		$\pi^* = 0.50$	0.501	0.001			
(200,5)		$\beta_0 = 0.20$	0.197	-0.003	0.148	0.123	95.4
		$\beta_1 = 0.60$	0.592	-0.008	0.226	0.201	95.7
		$\beta_2 = 0.50$	0.503	0.003	0.146	0.138	96.0
		$\rho_1 = 0.567$	0.571	0.004			
		$\alpha_0 = -1.00$	-1.006	-0.006	0.193	0.315	95.3
		$\alpha_1 = 0.40$	0.408	0.008	0.289	0.412	94.4
	0.908	$\alpha_2 = 0.50$	0.505	0.005	0.181	0.258	95.6
		$\rho_1^* = 0.16$	0.107	-0.053			
		$\pi^* = 0.50$	0.501	0.001			
(500,3)		$\beta_0 = 0.20$	0.200	0.000	0.093	0.080	96.3
		$\beta_1 = 0.60$	0.597	-0.003	0.153	0.137	95.8
		$\beta_2 = 0.50$	0.500	-0.000	0.098	0.089	95.7
		$\rho_1 = 0.567$	0.566	-0.001			
		$\alpha_0 = -1.00$	-1.064	-0.064	0.131	0.243	93.5
		$\alpha_1 = 0.40$	0.498	0.098	0.211	0.319	94.1
	0.910	$\alpha_2 = 0.50$	0.542	0.042	0.130	0.197	94.8
		$\rho_1^* = 0.16$	0.044	-0.116			
		$\pi^* = 0.50$	0.500	0.000			
(500,4)		$\beta_0 = 0.20$	0.195	-0.005	0.090	0.072	95.6
		$\beta_1 = 0.60$	0.608	0.008	0.149	0.122	95.5
		$\beta_2 = 0.50$	0.500	0.000	0.095	0.079	95.9
		$\rho_1 = 0.567$	0.564	-0.003			
		$\alpha_0 = -1.00$	-1.029	-0.029	0.121	0.195	95.0
		$\alpha_1 = 0.40$	0.440	0.040	0.197	0.273	94.8
	0.885	$\alpha_2 = 0.50$	0.520	0.020	0.120	0.170	94.9
		$\rho_1^* = 0.16$	0.083	-0.077			
		$\pi^* = 0.50$	0.500	0.000			
(500,5)		$\beta_0 = 0.20$	0.201	0.001	0.089	0.072	95.5
		$\beta_1 = 0.60$	0.598	-0.002	0.147	0.126	95.3
		$\beta_2 = 0.50$	0.496	-0.004	0.094	0.080	95.4
		$\rho_1 = 0.567$	0.566	-0.001			
		$\alpha_0 = -1.00$	-1.020	-0.020	0.116	0.181	94.9
		$\alpha_1 = 0.40$	0.428	0.028	0.188	0.250	95.200
	0.889	$\alpha_2 = 0.50$	0.517	0.017	0.115	0.155	93.8
		$\rho_1^* = 0.16$	0.107	-0.053			
		$\pi^* = 0.50$	0.499	-0.001			

Table C.6: Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.90$ from marginalized mixture of Poisson EQCOR model for $\rho^* = 0.40$ and for different values of K and T

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(100,3)		$\beta_0 = 0.20$	0.184	-0.016	0.503	0.565	95.2
		$\beta_1 = 0.60$	0.564	-0.036	0.788	0.921	93.8
		$\beta_2 = 0.50$	0.417	-0.083	0.553	0.624	93.7
		$\rho_1 = 0.909$	0.881	-0.028			
		$\alpha_0 = -1.00$	-1.000	-0.000	0.250	0.257	95.8
		$\alpha_1 = 0.40$	0.375	-0.025	0.372	0.384	95.2
	0.811	$\alpha_2 = 0.50$	0.484	-0.016	0.252	0.259	95.1
		$\rho_1^* = 0.16$	0.142	-0.018			
		$\pi^* = 0.90$	0.896	-0.004			
(100,4)		$\beta_0 = 0.20$	0.164	-0.036	0.496	0.551	95.9
		$\beta_1 = 0.60$	0.605	0.005	0.777	0.892	95.0
		$\beta_2 = 0.50$	0.443	-0.057	0.546	0.646	94.9
		$\rho_1 = 0.909$	0.880	-0.029			
		$\alpha_0 = -1.00$	-1.001	-0.001	0.229	0.235	94.9
		$\alpha_1 = 0.40$	0.389	-0.011	0.340	0.347	94.6
	0.820	$\alpha_2 = 0.50$	0.499	-0.001	0.230	0.235	96.1
		$\rho_1^* = 0.16$	0.148	-0.012			
		$\pi^* = 0.90$	0.896	-0.004			
(100,5)		$\beta_0 = 0.20$	0.113	-0.087	0.493	0.567	94.7
		$\beta_1 = 0.60$	0.678	0.078	0.777	0.934	93.8
		$\beta_2 = 0.50$	0.460	-0.040	0.546	0.625	93.7
		$\rho_1 = 0.909$	0.874	-0.035			
		$\alpha_0 = -1.00$	-0.996	0.004	0.212	0.209	94.9
		$\alpha_1 = 0.40$	0.392	-0.008	0.315	0.308	94.8
	0.808	$\alpha_2 = 0.50$	0.486	-0.014	0.214	0.229	95.2
		$\rho_1^* = 0.16$	0.142	-0.018			
		$\pi^* = 0.90$	0.897	-0.003			
(200,3)		$\beta_0 = 0.20$	0.162	-0.038	0.368	0.335	95.2
		$\beta_1 = 0.60$	0.606	0.006	0.584	0.564	95.5
		$\beta_2 = 0.50$	0.494	-0.006	0.383	0.378	94.6
		$\rho_1 = 0.909$	0.902	-0.007			
		$\alpha_0 = -1.00$	-1.008	-0.008	0.179	0.184	94.8
		$\alpha_1 = 0.40$	0.400	0.000	0.269	0.275	93.9
	0.851	$\alpha_2 = 0.50$	0.504	0.004	0.168	0.174	94.6
		$\rho_1^* = 0.16$	0.149	-0.011			
		$\pi^* = 0.90$	0.900	-0.000			
(200,4)		$\beta_0 = 0.20$	0.156	-0.044	0.363	0.336	95.3
		$\beta_1 = 0.60$	0.636	0.036	0.576	0.522	96.2
		$\beta_2 = 0.50$	0.519	0.019	0.379	0.363	94.9
		$\rho_1 = 0.909$	0.905	-0.004			
		$\alpha_0 = -1.00$	-1.018	-0.018	0.165	0.169	95.0
		$\alpha_1 = 0.40$	0.413	0.013	0.247	0.253	95.6
	0.847	$\alpha_2 = 0.50$	0.508	0.008	0.154	0.155	95.0
		$\rho_1^* = 0.16$	0.150	-0.010			
		$\pi^* = 0.90$	0.897	-0.003			

Continued...Table C.6

(K,T)	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
(200,5)		$\beta_0 = 0.20$	0.174	-0.026	0.366	0.340	95.0
		$\beta_1 = 0.60$	0.597	-0.003	0.578	0.560	94.5
		$\beta_2 = 0.50$	0.505	0.005	0.379	0.376	95.4
		$\rho_1 = 0.909$	0.905	-0.004			
		$\alpha_0 = -1.00$	-1.009	-0.009	0.155	0.156	95.1
		$\alpha_1 = 0.40$	0.403	0.003	0.232	0.230	94.7
	0.812	$\alpha_2 = 0.50$	0.504	0.004	0.145	0.148	95.8
		$\rho_1^* = 0.16$	0.155	-0.005			
		$\pi^* = 0.90$	0.898	-0.002			
(500,3)		$\beta_0 = 0.20$	0.180	-0.020	0.223	0.191	96.2
		$\beta_1 = 0.60$	0.609	0.009	0.383	0.339	96.2
		$\beta_2 = 0.50$	0.496	-0.004	0.249	0.222	95.5
		$\rho_1 = 0.909$	0.907	-0.002			
		$\alpha_0 = -1.00$	-1.001	-0.001	0.108	0.108	95.1
		$\alpha_1 = 0.40$	0.397	-0.003	0.175	0.179	94.2
	0.880	$\alpha_2 = 0.50$	0.499	-0.001	0.107	0.105	94.3
		$\rho_1^* = 0.16$	0.158	-0.002			
		$\pi^* = 0.90$	0.900	0.000			
(500,4)		$\beta_0 = 0.20$	0.174	-0.026	0.221	0.198	95.1
		$\beta_1 = 0.60$	0.628	0.028	0.381	0.338	95.8
		$\beta_2 = 0.50$	0.512	0.012	0.248	0.212	95.8
		$\rho_1 = 0.909$	0.907	-0.002			
		$\alpha_0 = -1.00$	-1.002	-0.002	0.099	0.097	95.4
		$\alpha_1 = 0.40$	0.396	-0.004	0.160	0.162	96.3
	0.886	$\alpha_2 = 0.50$	0.500	-0.000	0.098	0.099	95.5
		$\rho_1^* = 0.16$	0.157	-0.003			
		$\pi^* = 0.90$	0.899	-0.001			
(500,5)		$\beta_0 = 0.20$	0.190	-0.010	0.221	0.191	96.0
		$\beta_1 = 0.60$	0.588	-0.012	0.380	0.341	95.9
		$\beta_2 = 0.50$	0.499	-0.001	0.247	0.222	96.0
		$\rho_1 = 0.909$	0.907	-0.002			
		$\alpha_0 = -1.00$	-0.999	0.001	0.093	0.093	95.6
		$\alpha_1 = 0.40$	0.395	-0.005	0.151	0.154	94.8
	0.883	$\alpha_2 = 0.50$	0.500	-0.000	0.092	0.091	94.9
		$\rho_1^* = 0.16$	0.157	-0.003			
		$\pi^* = 0.90$	0.900	-0.000			

C.4 Simulation Results: Marginalized Mixture of Longitudinal Poisson Model

Table C.7: Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.50$ from marginalized mixture of Poisson AR(1), Poisson MA(1), and Poisson EQCOR model for $\rho^* = 0.70$ and $T = 4$, and for different values of K

Mixture Type	K	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
AR(1)	100		$\beta_0 = 0.20$	0.175	-0.025	0.237	0.208	95.8
			$\beta_1 = 0.60$	0.611	0.011	0.356	0.308	94.9
			$\beta_2 = 0.50$	0.499	-0.001	0.249	0.227	96.0
		0.905	$\rho_1 = 0.848$	0.839	-0.009			
			$\alpha_0 = -1.00$	-1.412	-0.412	0.413	0.964	93.7
			$\alpha_1 = 0.40$	0.687	0.287	0.595	1.123	95.4
			$\alpha_2 = 0.50$	0.711	0.211	0.403	0.733	94.0
			$\pi^* = 0.50$	0.498	-0.002			
	200		$\beta_0 = 0.20$	0.205	0.005	0.169	0.144	95.4
			$\beta_1 = 0.60$	0.585	-0.015	0.257	0.223	96.1
			$\beta_2 = 0.50$	0.499	-0.001	0.166	0.146	95.8
		0.908	$\rho_1 = 0.848$	0.845	-0.003			
			$\alpha_0 = -1.00$	-1.229	-0.229	0.279	0.500	92.0
			$\alpha_1 = 0.40$	0.549	0.149	0.412	0.654	94.3
			$\alpha_2 = 0.50$	0.622	0.122	0.256	0.394	93.7
			$\pi^* = 0.50$	0.499	-0.001			
	500		$\beta_0 = 0.20$	0.202	0.002	0.101	0.084	95.9
			$\beta_1 = 0.60$	0.594	-0.006	0.167	0.141	95.3
			$\beta_2 = 0.50$	0.500	-0.000	0.106	0.090	96.1
		0.900	$\rho_1 = 0.848$	0.844	-0.004			
			$\alpha_0 = -1.00$	-1.217	-0.217	0.163	0.327	89.3
			$\alpha_1 = 0.40$	0.579	0.179	0.260	0.401	92.0
			$\alpha_2 = 0.50$	0.608	0.108	0.159	0.255	92.2
			$\pi^* = 0.50$	0.499	-0.001			
MA(1)	100		$\beta_0 = 0.20$	0.193	-0.007	0.212	0.177	96.2
			$\beta_1 = 0.60$	0.596	-0.004	0.321	0.280	95.3
			$\beta_2 = 0.50$	0.506	0.006	0.224	0.193	95.8
		0.896	$\rho_1 = 0.703$	0.694	-0.009			
			$\alpha_0 = -1.00$	-1.072	-0.072	0.300	0.573	95.0
			$\alpha_1 = 0.40$	0.460	0.060	0.444	0.709	95.8
			$\alpha_2 = 0.50$	0.547	0.047	0.302	0.480	96.0
			$\pi^* = 0.50$	0.499	-0.001			
	200		$\beta_0 = 0.20$	0.186	-0.014	0.152	0.132	96.2
			$\beta_1 = 0.60$	0.610	0.010	0.232	0.209	96.2
			$\beta_2 = 0.50$	0.510	0.010	0.150	0.129	95.6
		0.905	$\rho_1 = 0.703$	0.699	-0.004			
			$\alpha_0 = -1.00$	-1.055	-0.055	0.214	0.374	96.0

Continued...Table C.7

Mixture Type	K	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
			$\alpha_1 = 0.40$	0.466	0.066	0.320	0.481	95.5
			$\alpha_2 = 0.50$	0.548	0.048	0.200	0.309	94.8
			$\pi^* = 0.50$	0.501	0.001			
	500		$\beta_0 = 0.20$	0.197	-0.003	0.091	0.076	96.0
			$\beta_1 = 0.60$	0.602	0.002	0.150	0.134	95.3
			$\beta_2 = 0.50$	0.501	0.001	0.096	0.084	95.7
		0.908	$\rho_1 = 0.703$	0.696	-0.007			
			$\alpha_0 = -1.00$	-1.043	-0.043	0.127	0.207	94.9
			$\alpha_1 = 0.40$	0.452	0.052	0.206	0.288	95.2
			$\alpha_2 = 0.50$	0.528	0.028	0.126	0.173	94.8
			$\pi^* = 0.50$	0.499	-0.001			
EQCOR	100		$\beta_0 = 0.20$	0.183	-0.017	0.233	0.198	96.0
			$\beta_1 = 0.60$	0.612	0.012	0.352	0.313	95.5
			$\beta_2 = 0.50$	0.489	-0.011	0.245	0.226	96.0
		0.915	$\rho_1 = 0.743$	0.730	-0.013			
			$\alpha_0 = -1.00$	-1.323	-0.323	0.383	0.817	93.8
			$\alpha_1 = 0.40$	0.606	0.206	0.557	0.969	95.3
			$\alpha_2 = 0.50$	0.651	0.151	0.378	0.677	94.9
			$\pi^* = 0.50$	0.498	-0.002			
	200		$\beta_0 = 0.20$	0.200	0.000	0.167	0.141	96.9
			$\beta_1 = 0.60$	0.591	-0.009	0.254	0.220	95.3
			$\beta_2 = 0.50$	0.493	-0.007	0.164	0.138	95.4
		0.907	$\rho_1 = 0.743$	0.737	-0.006			
			$\alpha_0 = -1.00$	-1.230	-0.230	0.268	0.512	93.9
			$\alpha_1 = 0.40$	0.553	0.153	0.396	0.621	95.3
			$\alpha_2 = 0.50$	0.610	0.110	0.247	0.382	94.5
			$\pi^* = 0.50$	0.499	-0.001			
	500		$\beta_0 = 0.20$	0.197	-0.003	0.100	0.084	96.4
			$\beta_1 = 0.60$	0.601	0.001	0.165	0.143	95.1
			$\beta_2 = 0.50$	0.499	-0.001	0.105	0.090	96.7
		0.917	$\rho_1 = 0.743$	0.735	-0.008			
			$\alpha_0 = -1.00$	-1.183	-0.183	0.156	0.294	90.4
			$\alpha_1 = 0.40$	0.536	0.136	0.250	0.389	92.5
			$\alpha_2 = 0.50$	0.596	0.096	0.152	0.243	93.1
			$\pi^* = 0.50$	0.501	0.001			

Table C.8: Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.70$ from marginalized mixture of Poisson AR(1), Poisson MA(1), and Poisson EQCOR model for $\rho^* = 0.70$ and $T = 4$, and for different values of K

Mixture Type	K	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
AR(1)	100		$\beta_0 = 0.20$	0.166	-0.034	0.298	0.277	95.2
			$\beta_1 = 0.60$	0.627	0.027	0.461	0.443	95.8
			$\beta_2 = 0.50$	0.491	-0.009	0.324	0.301	95.1
		0.880	$\rho_1 = 0.905$	0.901	-0.004			
			$\alpha_0 = -1.00$	-0.982	0.018	0.333	0.443	93.9
			$\alpha_1 = 0.40$	0.365	-0.035	0.499	0.633	94.9
			$\alpha_2 = 0.50$	0.457	-0.043	0.341	0.436	94.9
			$\pi^* = 0.70$	0.693	-0.007			
	200		$\beta_0 = 0.20$	0.177	-0.023	0.213	0.192	95.2
			$\beta_1 = 0.60$	0.624	0.024	0.331	0.295	95.8
			$\beta_2 = 0.50$	0.510	0.010	0.216	0.195	94.8
		0.910	$\rho_1 = 0.905$	0.904	-0.001			
			$\alpha_0 = -1.00$	-0.974	0.026	0.237	0.324	94.8
			$\alpha_1 = 0.40$	0.392	-0.008	0.358	0.460	95.3
			$\alpha_2 = 0.50$	0.480	-0.020	0.225	0.292	95.1
			$\pi^* = 0.70$	0.695	-0.005			
	500		$\beta_0 = 0.20$	0.191	-0.009	0.129	0.109	96.1
			$\beta_1 = 0.60$	0.605	0.005	0.217	0.187	95.3
			$\beta_2 = 0.50$	0.505	0.005	0.140	0.120	96.0
		0.908	$\rho_1 = 0.905$	0.903	-0.002			
			$\alpha_0 = -1.00$	-0.959	0.041	0.141	0.185	94.4
			$\alpha_1 = 0.40$	0.375	-0.025	0.231	0.281	94.4
			$\alpha_2 = 0.50$	0.478	-0.022	0.142	0.170	94.4
			$\pi^* = 0.70$	0.700	-0.000			
MA(1)	100		$\beta_0 = 0.20$	0.168	-0.032	0.285	0.246	94.8
			$\beta_1 = 0.60$	0.614	0.014	0.442	0.396	95.3
			$\beta_2 = 0.50$	0.490	-0.010	0.310	0.275	96.3
		0.886	$\rho_1 = 0.814$	0.810	-0.004			
			$\alpha_0 = -1.00$	-0.931	0.069	0.262	0.327	94.1
			$\alpha_1 = 0.40$	0.326	-0.074	0.394	0.472	93.8
			$\alpha_2 = 0.50$	0.454	-0.046	0.267	0.323	95.5
			$\pi^* = 0.70$	0.699	-0.001			
	200		$\beta_0 = 0.20$	0.197	-0.003	0.202	0.166	96.4
			$\beta_1 = 0.60$	0.596	-0.004	0.316	0.259	95.7
			$\beta_2 = 0.50$	0.492	-0.008	0.206	0.183	95.6
		0.879	$\rho_1 = 0.814$	0.817	0.003			
			$\alpha_0 = -1.00$	-0.923	0.077	0.188	0.245	92.5
			$\alpha_1 = 0.40$	0.326	-0.074	0.285	0.337	94.7
			$\alpha_2 = 0.50$	0.448	-0.052	0.178	0.210	92.9
			$\pi^* = 0.70$	0.697	-0.003			
	500		$\beta_0 = 0.20$	0.202	0.002	0.122	0.105	95.8
			$\beta_1 = 0.60$	0.590	-0.010	0.206	0.179	95.0
			$\beta_2 = 0.50$	0.492	-0.008	0.133	0.120	95.7

Continued...Table C.8

Mixture Type	K	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
		0.904	$\rho_1 = 0.814$	0.813	-0.001			
			$\alpha_0 = -1.00$	-0.922	0.078	0.113	0.142	91.8
			$\alpha_1 = 0.40$	0.328	-0.072	0.186	0.218	93.6
			$\alpha_2 = 0.50$	0.454	-0.046	0.114	0.130	94.0
			$\pi^* = 0.70$	0.699	-0.001			
EQCOR	100		$\beta_0 = 0.20$	0.183	-0.017	0.298	0.255	96.0
			$\beta_1 = 0.60$	0.598	-0.002	0.460	0.408	96.2
			$\beta_2 = 0.50$	0.488	-0.012	0.323	0.287	96.1
		0.898	$\rho_1 = 0.839$	0.832	-0.007			
			$\alpha_0 = -1.00$	-0.978	0.022	0.323	0.436	94.8
			$\alpha_1 = 0.40$	0.360	-0.040	0.483	0.611	95.5
			$\alpha_2 = 0.50$	0.464	-0.036	0.331	0.431	94.5
			$\pi^* = 0.70$	0.696	-0.004			
	200		$\beta_0 = 0.20$	0.197	-0.003	0.213	0.174	96.0
			$\beta_1 = 0.60$	0.588	-0.012	0.332	0.288	95.3
			$\beta_2 = 0.50$	0.494	-0.006	0.216	0.198	95.4
		0.894	$\rho_1 = 0.839$	0.839	0.000			
			$\alpha_0 = -1.00$	-0.975	0.025	0.231	0.310	95.0
			$\alpha_1 = 0.40$	0.370	-0.030	0.348	0.438	95.5
			$\alpha_2 = 0.50$	0.475	-0.025	0.219	0.271	94.2
			$\pi^* = 0.70$	0.697	-0.003			
	500		$\beta_0 = 0.20$	0.193	-0.007	0.128	0.111	96.1
			$\beta_1 = 0.60$	0.597	-0.003	0.216	0.189	95.0
			$\beta_2 = 0.50$	0.507	0.007	0.139	0.123	95.0
		0.905	$\rho_1 = 0.839$	0.838	-0.001			
			$\alpha_0 = -1.00$	-0.948	0.052	0.137	0.172	94.5
			$\alpha_1 = 0.40$	0.344	-0.056	0.225	0.275	94.3
			$\alpha_2 = 0.50$	0.459	-0.041	0.138	0.157	93.4
			$\pi^* = 0.70$	0.699	-0.001			

Table C.9: Simulated mean (SM), amount of bias (Bias), estimated and simulated standard error (ESE, SSE) and coverage probability (Cov.Pr) in estimating marginal parameters ($\boldsymbol{\beta}$); and component-1 parameters ($\boldsymbol{\alpha}$) with mixing proportion $\pi^* = 0.90$ from marginalized mixture of Poisson AR(1), Poisson MA(1), and Poisson EQCOR model for $\rho^* = 0.70$ and $T = 4$, and for different values of K

Mixture Type	K	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
AR(1)	100		$\beta_0 = 0.20$	0.179	-0.021	0.515	0.590	93.8
			$\beta_1 = 0.60$	0.564	-0.036	0.802	0.967	94.4
			$\beta_2 = 0.50$	0.467	-0.033	0.567	0.646	94.5
		0.819	$\rho_1 = 0.967$	0.952	-0.015			
			$\alpha_0 = -1.00$	-1.041	-0.041	0.308	0.319	95.0
			$\alpha_1 = 0.40$	0.430	0.030	0.456	0.480	94.6
			$\alpha_2 = 0.50$	0.505	0.005	0.310	0.314	94.3
			$\pi^* = 0.90$	0.896	-0.004			
	200		$\beta_0 = 0.20$	0.168	-0.032	0.372	0.345	95.2
			$\beta_1 = 0.60$	0.614	0.014	0.590	0.561	94.8
			$\beta_2 = 0.50$	0.519	0.019	0.389	0.335	94.5
		0.806	$\rho_1 = 0.967$	0.964	-0.003			
			$\alpha_0 = -1.00$	-1.011	-0.011	0.220	0.232	95.3
			$\alpha_1 = 0.40$	0.402	0.002	0.330	0.351	95.2
			$\alpha_2 = 0.50$	0.500	-0.000	0.206	0.211	95.2
			$\pi^* = 0.90$	0.899	-0.001			
	500		$\beta_0 = 0.20$	0.171	-0.029	0.224	0.207	95.4
			$\beta_1 = 0.60$	0.633	0.033	0.386	0.337	96.3
			$\beta_2 = 0.50$	0.523	0.023	0.252	0.244	96.3
		0.890	$\rho_1 = 0.967$	0.966	-0.001			
			$\alpha_0 = -1.00$	-1.008	-0.008	0.132	0.137	94.7
			$\alpha_1 = 0.40$	0.410	0.010	0.213	0.221	94.4
			$\alpha_2 = 0.50$	0.504	0.004	0.131	0.139	95.1
			$\pi^* = 0.90$	0.899	-0.001			
MA(1)	100		$\beta_0 = 0.20$	0.151	-0.049	0.497	0.602	94.5
			$\beta_1 = 0.60$	0.596	-0.004	0.776	1.033	95.2
			$\beta_2 = 0.50$	0.421	-0.079	0.544	0.657	95.1
		0.830	$\rho_1 = 0.936$	0.903	0.033			
			$\alpha_0 = -1.00$	-1.011	-0.011	0.236	0.242	94.9
			$\alpha_1 = 0.40$	0.403	0.003	0.350	0.362	94.7
			$\alpha_2 = 0.50$	0.504	0.004	0.237	0.232	95.7
			$\pi^* = 0.90$	0.898	-0.002			
	200		$\beta_0 = 0.20$	0.171	-0.029	0.364	0.338	94.9
			$\beta_1 = 0.60$	0.592	-0.008	0.577	0.570	95.7
			$\beta_2 = 0.50$	0.507	0.007	0.380	0.350	95.5
		0.831	$\rho_1 = 0.936$	0.931	-0.005			
			$\alpha_0 = -1.00$	-1.002	-0.002	0.169	0.183	95.3
			$\alpha_1 = 0.40$	0.390	-0.010	0.254	0.273	95.5
			$\alpha_2 = 0.50$	0.498	-0.002	0.158	0.169	95.2
			$\pi^* = 0.90$	0.899	-0.001			
	500		$\beta_0 = 0.20$	0.177	-0.023	0.221	0.202	94.8
			$\beta_1 = 0.60$	0.611	0.011	0.382	0.346	95.5
			$\beta_2 = 0.50$	0.514	0.014	0.248	0.213	95.6

Continued...Table C.9

Mixture Type	K	Conv.Prop.	Params	SM	Bias	ESE	SSE	Cov.Pr.
		0.885	$\rho_1 = 0.936$	0.935	-0.001			
			$\alpha_0 = -1.00$	-0.998	0.002	0.102	0.105	94.8
			$\alpha_1 = 0.40$	0.392	-0.008	0.165	0.167	95.5
			$\alpha_2 = 0.50$	0.501	0.001	0.101	0.103	95.0
			$\pi^* = 0.90$	0.900	-0.000			
EQCOR	100		$\beta_0 = 0.20$	0.199	-0.001	0.508	0.595	95.7
			$\beta_1 = 0.60$	0.504	-0.096	0.792	1.004	94.8
			$\beta_2 = 0.50$	0.439	-0.061	0.561	0.676	95.3
		0.830	$\rho_1 = 0.945$	0.919	0.026			
			$\alpha_0 = -1.00$	-1.001	-0.001	0.296	0.309	94.8
			$\alpha_1 = 0.40$	0.375	-0.025	0.441	0.460	94.9
			$\alpha_2 = 0.50$	0.485	-0.015	0.299	0.312	94.9
			$\pi^* = 0.90$	0.896	-0.004			
	200		$\beta_0 = 0.20$	0.186	-0.014	0.371	0.365	95.1
			$\beta_1 = 0.60$	0.588	-0.012	0.586	0.596	95.0
			$\beta_2 = 0.50$	0.489	-0.011	0.385	0.375	94.8
		0.854	$\rho_1 = 0.945$	0.941	-0.004			
			$\alpha_0 = -1.00$	-1.030	-0.030	0.215	0.227	94.6
			$\alpha_1 = 0.40$	0.428	0.028	0.321	0.333	94.8
			$\alpha_2 = 0.50$	0.512	0.012	0.201	0.211	95.7
			$\pi^* = 0.90$	0.898	-0.002			
	500		$\beta_0 = 0.20$	0.192	-0.008	0.225	0.175	94.5
			$\beta_1 = 0.60$	0.607	0.007	0.387	0.313	96.1
			$\beta_2 = 0.50$	0.490	-0.010	0.251	0.218	96.1
		0.854	$\rho_1 = 0.945$	0.945	0.000			
			$\alpha_0 = -1.00$	-1.009	-0.009	0.128	0.129	95.2
			$\alpha_1 = 0.40$	0.409	0.009	0.207	0.199	94.6
			$\alpha_2 = 0.50$	0.501	0.001	0.127	0.131	94.5
			$\pi^* = 0.90$	0.899	-0.001			

Bibliography

Abramowitz, M., Stegun, I. A., et al. (1964). *Handbook of mathematical functions*, volume 55. Dover New York.

Afolabi, O. B. Y. R. F. and Agbaje, A. S. (2018). Modelling excess zeros in count data with application to antenatal care utilisation. *International Journal of Statistics and Probability*, 7(3).

Albert, J. M., Wang, W., and Nelson, S. (2014). Estimating overall exposure effects for zero-inflated regression models with application to dental caries. *Statistical methods in medical research*, 23(3):257–278.

Alfò, M. and Maruotti, A. (2010). Two-part regression models for longitudinal zero-inflated count data. *Canadian Journal of Statistics*, 38(2):197–216.

Bekalo, D. B. and Kebede, D. T. (2021). Zero-inflated models for count data: an application to number of antenatal care service visits. *Annals of Data Science*, 8(4):683–708.

Benecha, H. K., Neelon, B., Divaris, K., and Preisser, J. S. (2017). Marginalized mixture models for count data from multiple source populations. *Journal of statistical distributions and applications*, 4(1):1–17.

Bhowmik, K. R., Das, S., and Islam, M. A. (2020). Modelling the number of antenatal care visits in Bangladesh to determine the risk factors for reduced antenatal care attendance. *PloS one*, 15(1):e0228215.

Bilgic, A. and Florkowski, W. J. (2007). Application of a hurdle negative binomial count data model to demand for bass fishing in the southeastern United States. *Journal of environmental management*, 83(4):478–490.

Böhning, D., Dietz, E., Schlattmann, P., Mendonca, L., and Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(2):195–209.

Buu, A., Li, R., Tan, X., and Zucker, R. A. (2012). Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. *Statistics in medicine*, 31(29):4074–4086.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. New York: Cambridge university press.

Crowder, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika*, 82(2):407–410.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Dobbie, M. J. and Welsh, A. H. (2001). Theory & Methods: Modelling Correlated Zero-inflated Count Data. *Australian & New Zealand Journal of Statistics*, 43(4):431–444.

Farewell, V., Long, D., Tom, B., Yiu, S., and Su, L. (2017). Two-part and related regression models for longitudinal data. *Annual review of statistics and its application*, 4:283–315.

Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, 80(1):141–151.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012). *Applied longitudinal analysis*, volume 998. John Wiley & Sons.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.

Golub, G. H. (1973). Some modified matrix eigenvalue problems. *SIAM review*, 15(2):318–334.

Greene, W. H. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models.

Gurmu, S. (1997). Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization. *Journal of applied econometrics*, 12(3):225–242.

Gurmu, S. and Trivedi, P. K. (1996). Excess zeros in count models for recreational trips. *Journal of Business & Economic Statistics*, 14(4):469–477.

Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, 56(4):1030–1039.

Hall, D. B. and Severini, T. A. (1998). Extended generalized estimating equations for clustered data. *Journal of the American Statistical Association*, 93(444):1365–1375.

Hall, D. B. and Zhang, Z. (2004). Marginal models for zero inflated clustered data. *Statistical modelling*, 4(3):161–180.

Haque, M. E., Mallick, T. S., and Bari, W. (2022a). Does dropout from school matter in taking antenatal care visits among women in Bangladesh? An application of marginalized poisson-poisson mixture model. *BMC Pregnancy and Childbirth*, 22(1):1–8.

Haque, M. E., Mallick, T. S., and Bari, W. (2022b). Zero truncated Poisson model: an alternative approach for analyzing count data with excess zeros. *Journal of Statistical Computation and Simulation*, 92(3):476–487.

Hasan, M. T., Huda, S., and Sneddon, G. (2016). A comparative study of observation-and parameter-driven zero-inflated Poisson models for longitudinal count data. *Communications in Statistics-Simulation and Computation*, 45(10):3643–3659.

Hasan, M. T. and Sneddon, G. (2009). Zero-inflated Poisson regression for longitudinal data. *Communications in Statistics—Simulation and Computation®*, 38(3):638–653.

Hilbe, J. M. (2014). *Modeling count data*. New York: Cambridge University Press.

Hossain, Z., Akter, R., Sultana, N., and Kabir, E. (2020). Modelling zero-truncated overdispersed antenatal health care count data of women in bangladesh. *PloS one*, 15(1):e0227824.

Hu, M.-C., Pavlicova, M., and Nunes, E. V. (2011). Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *The American journal of drug and alcohol abuse*, 37(5):367–375.

Islam, M. M. and Masud, M. S. (2018). Determinants of frequency and contents of antenatal care visits in Bangladesh: Assessing the extent of compliance with the WHO recommendations. *PloS one*, 13(9).

Kassahun, W., Neyens, T., Molenberghs, G., Faes, C., and Verbeke, G. (2014). Marginalized multilevel hurdle and zero-inflated models for overdispersed and correlated count data with excess zeros. *Statistics in medicine*, 33(25):4402–4419.

Killip, S., Mahfoud, Z., and Pearce, K. (2004). What is an intracluster correlation coefficient? crucial concepts for primary care researchers. *The Annals of Family Medicine*, 2(3):204–208.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14.

Lee, K., Joo, Y., Song, J. J., and Harper, D. W. (2011). Analysis of zero-inflated clustered count data: a marginalized model approach. *Computational Statistics & Data Analysis*, 55(1):824–837.

Leisch, F. (2004). Flexmix: A general framework for finite mixture models and latent glass regression in R.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Long, D. L., Preisser, J. S., Herring, A. H., and Golin, C. E. (2014). A marginalized zero-inflated Poisson regression model with overall exposure effects. *Statistics in medicine*, 33(29):5151–5165.

Long, D. L., Preisser, J. S., Herring, A. H., and Golin, C. E. (2015). A marginalized zero-inflated poisson regression model with random effects. *Journal of the Royal Statistical Society. Series C, Applied statistics*, 64(5):815.

McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics*, 11(1):59–67.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, 20(4):822–835.

McLachlan, G. J. and Peel, D. (2000). *Finite mixture models.* Wiley Series in Probability and Statistics, New York.

Mekonnen, F. H., Lakew, W. D., Tesfaye, Z. D., and Swain, P. K. (2019). Statistical models for longitudinal zero-inflated count data: application to seizure attacks. *African Health Sciences*, 19(3):2555–2564.

Min, Y. and Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical modelling*, 5(1):1–19.

Minami, M., Lennert-Cody, C. E., Gao, W., and Román-Verdesoto, M. (2007). Modeling shark bycatch: the zero-inflated negative binomial regression model with smoothing. *Fisheries Research*, 84(2):210–221.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of econometrics*, 33(3):341–365.

Murray, D. M. and Short, B. J. (1997). Intraclass correlation among measures related to tobacco use by adolescents: estimates, correlates, and applications in intervention studies. *Addictive behaviors*, 22(1):1–12.

Oliveira, I. R., Molenberghs, G., Demétrio, C. G., Dias, C. T., Giolo, S. R., and Andrade, M. C. (2016). Quantifying intraclass correlations for count and time-to-event data. *Biometrical Journal*, 58(4):852–867.

Peerawaranun, P., Landier, J., Nosten, F. H., Nguyen, T.-N., Hien, T. T., Tripura, R., Peto, T. J., Phommasone, K., Mayxay, M., Day, N. P., et al. (2019). Intracluster correlation coefficients in the greater mekong subregion for sample size calculations of cluster randomized malaria trials. *Malaria Journal*, 18:1–9.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, 4(1):12–35.

Pohlmeier, W. (1996). An Econometric Model of the Two-Part Decision Making Process in the Demand for Health Care. *Journal of Human Resources*, 27(1):99–123.

Porter, M. D., White, G., et al. (2012). Self-exciting hurdle models for terrorist activity. *The Annals of Applied Statistics*, 6(1):106–124.

Preisser, J. S., Stamm, J. W., Long, D. L., and Kincade, M. E. (2012). Review and recommendations for zero-inflated count regression modeling of dental caries indices in epidemiological studies. *Caries research*, 46(4):413–423.

Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, pages 825–839.

Ridout, M. S., Demetrio, C. G., and Firth, D. (1999). Estimating intraclass correlation for binary data. *Biometrics*, 55(1):137–148.

Rose, C. E., Martin, S. W., Wannemuehler, K. A., and Plikaytis, B. D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *Journal of biopharmaceutical statistics*, 16(4):463–481.

Shankar, V., Milton, J., and Mannering, F. (1997). Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. *Accident Analysis & Prevention*, 29(6):829–837.

Shonkwiler, J. S. and Shaw, W. D. (1996). Hurdle count-data models in recreation demand analysis. *Journal of Agricultural and Resource Economics*, pages 210–219.

Su, L., Tom, B. D., and Farewell, V. T. (2015). A likelihood-based two-part marginal model for longitudinal semicontinuous data. *Statistical methods in medical research*, 24(2):194–205.

Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statistical Science*, 18(3):377–393.

Sutradhar, B. C. (2011). *Dynamic mixed models for familial longitudinal data*, volume 18. Springer.

Sutradhar, B. C. and Bari, W. (2007). On generalized quasilikelihood inference in longitudinal mixed model for count data. *Sankhyā: The Indian Journal of Statistics*, pages 671–699.

Sutradhar, B. C. and Das, K. (1999). Miscellanea. on the efficiency of regression estimators in generalised linear models for longitudinal data. *Biometrika*, 86(2):459–465.

Sutradhar, B. C. and Kovacevic, M. (2000). Analysing ordinal longitudinal survey data: Generalised estimating equations approach. *Biometrika*, 87(4):837–848.

Sutradhar, B. C. and Kumar, P. (2001). On the efficiency of extended generalized estimating equation approaches. *Statistics & probability letters*, 55(1):53–61.

Tabb, L. P., Tchetgen, E. J. T., Wellenius, G. A., and Coull, B. A. (2016). Marginalized zero-altered models for longitudinal count data. *Statistics in biosciences*, 8(2):181–203.

Tom, B. D., Su, L., and Farewell, V. T. (2016). A corrected formulation for marginal inference derived from two-part mixed models for longitudinal semi-continuous data. *Statistical methods in medical research*, 25(5):2014–2020.

Wang, K., Yau, K., and Lee, A. (2002). A hierarchical poisson mixture regression model to analyse maternity length of hospital stay. *Statistics in medicine*, 21(23):3639–3654.

Wang, K., Yau, K., Lee, A., and McLachlan, G. (2007). Two-component poisson mixture regression modelling of count data with bivariate random effects. *Mathematical and Computer Modelling*, 46(11-12):1468–1476.

Wang, P. (1994). *Mixed regression models for discrete data*. PhD thesis, University of British Columbia.

Wang, P., Puterman, M. L., Cockburn, I., and Le, N. (1996). Mixed Poisson regression models with covariate dependent rates. *Biometrics*, pages 381–400.

Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method. *Biometrika*, 61(3):439–447.

Wedel, M. and DeSarbo, W. S. (1995). A mixture likelihood approach for generalized linear models. *Journal of classification*, 12(1):21–55.

Welsh, A. H., Cunningham, R. B., Donnelly, C., and Lindenmayer, D. B. (1996). Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*, 88(1-3):297–308.

Yau, K. K., Wang, K., and Lee, A. H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal: journal of mathematical methods in biosciences*, 45(4):437–452.

Yip, K. C. and Yau, K. K. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, 36(2):153–163.

Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika*, 77(3):642–648.

Zhao, L. P., Prentice, R. L., and Self, S. G. (1992). Multivariate mean parameter estimation by using a partly exponential model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3):805–811.

Zhu, H., Luo, S., and DeSantis, S. M. (2017). Zero-inflated count models for longitudinal measurements with heterogeneous random effects. *Statistical methods in medical research*, 26(4):1774–1786.